OPEN ACCESS

University of BRISTOL

Nagy, C., Suderman, M. J., & al., E. (2020). Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature Neuroscience*, *2020*. https://doi.org/10.1038/s41593-020-0621-y

Peer reviewed version

Link to published version (if available):
10.1038/s41593-020-0621-y

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Nature Research at https://www.nature.com/articles/s41593-020-0621-y . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates OPCs and excitatory neurons

Corina Nagy[#,1], Malosree Maitra[#,1], Arnaud Tanti[1], Matthew Suderman[5], Jean-Francois Théroux[1], Maria Antonietta Davoli[1], Kelly Perlman[1], Volodymyr Yerko[1], Yu Chang Wang[3,4], Shreejoy J. Tripathy[6,7], Paul Pavlidis[6], Naguib Mechawar[1,2], Jiannis Ragoussis[3,4], Gustavo Turecki[1,2,3]*

[1]McGill Group for Suicide Studies, Douglas Mental Health University Institute, Montreal, Qc, Canada. [2]Department of Psychiatry, McGill University, Montreal, Qc, Canada. [3]Department of Human Genetics, McGill University, Montreal, Qc, Canada. [4]McGill University and Genome Quebec Innovation Centre, Montreal, Qc, Canada. [5]MRC Integrative Epidemiology Unit, University of Bristol, UK. [6]Michael Smith Laboratories and Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada. [7]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada. [8]Department of Psychiatry, University of Toronto, Toronto ON, Canada.

[#] These authors contributed equally

***Correspondence to**: Dr. Gustavo Turecki (gustavo.turecki@mcgill.ca)
6875 LaSalle Blvd, Verdun, Québec, Canada H4H 1R3
Tel.: (514) 761-6131 (ext. 2370)

Abstract: 142

Words: 4,462

Figures: 6

Tables:1

Major depressive disorder (MDD) has an enormous impact on global disease burden, affecting millions of people worldwide and ranking as a leading cause of disability for almost three decades. Past molecular studies of MDD employed bulk homogenates of post-mortem brain tissue, obscuring gene expression changes within individual cell types. Here, we used single-nucleus transcriptomics to examine ~80,000 nuclei from the dorsolateral prefrontal cortex of male individuals with MDD (n=17) and healthy controls (n=17). We identified 26 cellular clusters, and over 60% of these showed differential gene expression between groups. We found the greatest dysregulation in deep layer excitatory neurons and immature oligodendrocyte precursor cells (OPCs), contributing almost half (47%) of all changes in gene expression. These results highlight the importance of dissecting cell-type specific contributions to the disease, and offer opportunities to identify new avenues of research and novel targets for treatment.

Major depressive disorder (MDD) is a complex and heterogeneous disorder that affects an estimated 300 million people worldwide[1]. Genetic factors underlying the risk for MDD have been investigated using including genome-wide association studies, among other approaches[2]. Although some genetic associations have been detected, it remains a challenge to extract causal disease mechanisms from these findings [3]. It has been positing that MDD results from dysregulation of monoaminergic transmission, largely implicating the serotonergic and noradrenergic systems, has dominated the field for several decades. More recently, other factors have been associated with MDD, including glutamatergic and GABAergic transmission[4, 5], glial cell function, including astrocytic and oligodendrocytic contributions[6-8], blood-brain barrier integrity[6], and inflammation[9]. Given the wide variety of cell types in the brain and their complex interactions, investigative approaches with cell-type specificity are especially needed to gain insight into psychiatric phenotypes including MDD.

The interpretation of differential gene expression in bulk brain tissue homogenates is complicated by the heterogeneous cellular composition of the sample. Single-cell sequencing approaches have revealed that gene expression patterns in the brain are cell type specific, not only differentiating major classes of cells such as neuronal and glial cells, but even differentiating subtypes of glial cells and neurons[10, 11]. Therefore, it is difficult to verify whether subtle molecular differences observed from tissue homogenates are explained by the disease state or by differences in cell type composition between samples[12] Recently developed techniques for high-throughput single-cell and single-nucleus RNA-sequencing provide a solution for addressing this inherent drawback to bulk tissue experiments[11, 13].

High-throughput droplet-based single-nucleus RNA-sequencing (snRNA-seq) allows the profiling of thousands of nuclear transcriptomes, by utilizing nucleus-specific barcodes and unique

molecular identifiers (UMI) to tag individual RNA molecules. snRNA-seq yields comparable, albeit distinct, information[14] from single-cell RNA-seq (scRNA-seq), while facilitating the analysis of frozen tissues, which are not amenable to the isolation of intact cells. While there has been considerable interest in using scRNA-seq and snRNA-seq datasets to gain insight into the processes underlying complex brain disorders[15-17], very few direct comparisons of single-nucleus human brain gene expression has yet been performed in a psychiatric phenotype using high-throughput technologies.

Here, we sequenced ~80,000 nuclear transcriptomes from the prefrontal cortex of MDD cases and psychiatrically healthy controls and identified cell type specific differentially expressed genes. These results point to gene expression changes in predominantly two cell types, oligodendrocyte precursor cells and deep layer excitatory neurons. The relationships between and functions of the differentially expressed genes from these two cell clusters suggest impairments to FGF signalling, steroid hormone receptor cycling, immune function, and altered cytoskeletal regulation (related to changes in synaptic plasticity). This approach to snRNA-seq can effectively interrogate subtle phenotypes with improved resolution in archived brain tissue, and provide novel directions for follow-up studies.

Results

To assess the involvement of individual cell types in the pathophysiology of MDD, we examined nuclei from the dorsolateral prefrontal cortex (dlPFC), a region implicated in the pathology of major depressive disorder[18]. We used a droplet-based single-nucleus method optimized for use with postmortem brain tissue to assess a large number of nuclei. We measured 78,886 nuclei from 34 brain samples, half from patients who died during an episode of MDD, and the other

half from matched psychiatrically healthy individuals (Table 1, Supplementary Tables 1-3). The experimental design is depicted in Fig. 1. On average, we sequenced to a depth of almost 200 million reads per sample (Supplementary Table 1). Given that glial cells have consistently been found to have fewer transcripts than neuronal cells[10, 11], we used custom filtering criteria based on the distribution of UMIs per nucleus detected to recover a substantial number of glial cells (see Methods, Supplementary Fig. 1a-e, Supplementary Table 4). In an initial subset of 20 subjects, applying our custom filtering increased the total number of cells 1.8–fold but increased the number of non-neuronal cells by almost 6-fold (data not shown). More than 90% of the nuclei passing these filtering criteria had less than 5% reads from mitochondrially encoded genes (Supplementary Fig. 1f). The average gene count across nuclei ranged from 2144 in neurons to 1144 genes in glia (Supplementary Table 5). UMI counts were approximately twice the gene count for all cell types, as expected for this level of sequencing depth (Supplementary Table 5). Between sample groups, there were no significant differences between cases and controls in the median gene count per nucleus (t test p=0.12), median UMI count per nucleus (t test p=0.14), and number of cells detected per individual (t test=0.07) (Supplementary Table 1).

Identification of 26 distinct cell types in the dlPFC

In order to identify different cell types present in the brain samples, we applied unsupervised graph-based clustering[19] using the first 50 principal components derived from the 2135 most variable genes across individual nuclei (Methods, Supplementary Fig. 2a-b). After stringent quality control (Methods), we identified 26 distinct clusters (Fig. 2a). Each cluster was annotated using a combination of known cell type markers for excitatory and inhibitory neurons, and non-neuronal cells, including astrocytes, oligodendrocytes, oligodendrocyte precursor cells (OPCs), endothelial cells, and microglia (see Methods for full list of markers, Supplementary Table 6,

Supplementary Fig. 3a-p). Gene expression patterns specific to cell type clusters were visualised using a DotPlot (Fig. 2b), average and median gene expression heatmaps (Supplementary Fig. 4a-b), and violin plots (Fig. 2c-e) to form a consensus for annotation.

Refined cell subtypes reflect cortical cellular architecture

The clusters generated from our data are consistent with those previously reported in snRNA-seq of human PFC (Supplementary Fig. 5)[11]. Gene expression patterns previously linked to specific cortical layers (see Methods) coincide with our clustering of excitatory cells. In Fig. 2c, the genes are arranged from top to bottom in order of their expression across the cortical layers (first 17 rows, from the layer I/II to layer VI). There is a gradient of expression of these genes across the excitatory clusters. For example, clusters Ex1, Ex4, and Ex7-9 had high expression of *TLE4* (layer VI specific). Ex1, Ex8, and Ex9 showed concurrent expression of layer V/VI markers such as *TOX*. Ex6 and Ex7 additionally showed expression of the layer IV specific gene *RORB*. *HTR2C*, which is specific to a subset of layer V neurons, was prominent in Ex1 alone. *PCP4*, which is also layer V specific, was present in Ex1-3, Ex7, and Ex9. Superficial layer (I-III) markers such as *CUX2* and *RASGRF2* were mainly seen in the large cluster, Ex10. Likewise, inhibitory cell types demonstrated subtype specific gene expression patterns. For example, In7 was classified as inhibitory-parvalbumin because it expressed *GAD1* and *PVALB*, and lacked *VIP* and *SST* (Fig. 2d). Multiple astrocytic clusters were also identified, and while the typical sub-classification of astrocytes is based on their morphology within grey or white matter[20], we used only grey matter for these samples. As such, based on the higher percentage of *GFAP* expression in Astros_3 (38%) compared to Astros_2 (21%), we suspect that Astros_3 is more likely to represent reactive astrocytes[21] (Supplementary Table 6).

Reconstruction of oligodendrocyte developmental trajectory

We identified five distinct cell type clusters that fell into the oligodendrocyte lineage (OL), including two that we classified as OPCs (Fig 2e). OPCs express a characteristic set of markers such as *PDGFRA* and *PCDH15*, which decline as these cells mature into oligodendrocytes, whereas other lineage markers like*, OLIG2* or *SOX10*, are present in both mature and immature cells. Given these developmental stage specific markers it was possible to plot a pseudotime trajectory[22] using gene expression for OPC1, OPC2, Oligos1, Oligos2 and Oligos3. Our result indicated that OPC2 were the youngest cells within the dataset followed by OPC1, then Oligos2 and Oligos3, with Oligos1 being the most mature (Fig.2e, top). The expression of thousands of genes varied according to pseudotime ($q<0.01$). Approximately half of the genes associated with pseudotime overlapped in cases and controls (Supplementary Fig. 6a). However, among the genes exclusively associated with pseudotime in cases, there was a 2.7–fold enrichment of apoptosis signalling in PANTHER[23] pathway analysis (FDR $p<9.01\times10^{-3}$), while no enrichment was observed in controls. Given that certain stages of oligodendrocyte differentiation are associated with heightened susceptibility to apoptosis, this may indicate differences in OL development between cases and controls[24]. To assess the individual profiles of important developmental gene markers, we plotted their expression across pseudotime (Supplementary Fig. 6b-i), revealing their expected pattern of expression.

To compare our oligodendrocyte lineage (OL) cells with previously described OL cell types, we performed bioinformatic deconvolution (Fig. 2e, bottom). Our OPC2 gene expression profile was entirely represented by the "OPCs" gene expression profile from Jäkel et al. (2019)[25]. The OPC1 profile also primarily corresponded to the OPCs, but consistent with this cluster being further along the pseudotime trajectory, it showed a small correspondence to the COPs (committed oligodendrocyte precursors). Our oligodendrocyte clusters showed varying degrees of correspondence to the published data, with decreasing overlap to the published "OPCs"

expression profile with increasing maturity of the cell type (ranging from 70-11% correspondence). Interestingly, among our oligodendrocytes, Oligos3 showed the highest correspondence to the ImOlGs (immune oligodendroglia), as defined by Jäkel et al[25]. The "immune gene expression" feature of Oligos3 is highlighted in our hierarchical clustering dendrogram (Fig. 1b), in which Oligos3 is located closer to the Micro/Macro cluster compared to the other OL clusters.

Cell type-specific patterns of altered gene expression in MDD

We set out to assess gene expression differences between cases and controls within each cluster. However, one limitation of droplet based single-nucleus technology is the possibility of capturing doublet or multiplet nuclei, which we have estimated to be minimal in our case, as only 5.2% of captured nuclei were doublets or multiplets, based on a species mixing experiment (Supplementary Fig. 1g). This, however, represented a potential confounding factor when assessing differential gene expression between groups. We therefore eliminated doublets and multiplets from the dataset by calculating the correlation of each cell to the median expression value of its assigned cluster (Methods, Supplementary Fig. 7) and cells with low correlation were removed (Supplementary Table 7a-b). We also excluded any genes expressed in less than 10% of the cells in that cluster. Using only these purified clusters and filtered genes (median 5212 per cluster), we performed a differential gene expression analysis (Supplementary Tables 8-31).

A total of 96 genes (FDR <0.10) were differentially expressed in 16 of the 25 clusters analyzed (Fig. 3a) and 45 of those remained significant at FDR<0.05 (12 of 25 clusters). FDR correction considering all clusters together yields 41 significant genes (FDR < 0.10) in 16 clusters (Supplementary Table 32). This further supports that our statistical analyses are in fact able to detect differences in gene expression between the groups. To retain a larger set of genes in

order to better capture functional enrichments within individual cell types, we considered all genes which passed FDR < 0.10, corrected per cluster. The majority, 83% (80 genes), were downregulated in line with findings from previous transcriptomic studies in MDD[3, 4]. Differential expression analysis treated each cell as a sample (Supplementary Fig. 8a-f), but per subject contributions were visualized using heatmaps of average gene expression to assess biases in subject contributions. Patterns of gene expression averaged by subject reflected the expected differences between cases and controls (Supplementary Fig. 9a-p). Thirty-nine of the 96 differentially expressed genes were found in excitatory cell clusters and, of those, 34 were downregulated (Fig. 3a, insert). Some neuronal clusters contained both upregulated and downregulated genes, but it was more common for affected neuronal clusters to contain only downregulated genes (8/12, 67%). All but one inhibitory cluster showed altered gene expression and non-neuronal clusters tended to have both up- and downregulated genes (Fig. 3b).

Of particular interest, two clusters – one composed of immature oligodendrocyte precursor cells (OPC2) and one composed of deep layer excitatory neurons (Ex7) – accounted for almost half (47%) of the dysregulated genes (Fig. 3c). Finally, two genes were differentially expressed in more than one cluster: *PRKAR1B* showed decreased expression in excitatory clusters Ex7 (FDR=0.087, FC=0.87) and Ex2 (FDR=0.047, FC=0.82) and *TUBB4B* in excitatory clusters Ex7 (FDR=0.079, FC=0.87) and Ex6 (FDR=0.073, FC=0.86).

Cell type specific DEGs recapitulate published MDD findings

Three of our DEGs (*FADS2, CKB* and *KAZN*) have previously been identified in GWAS of MDD[2, 26].To further compare our DEGs with previously reported findings in MDD we took advantage of publically available databases PsyGeNET[27] and DisGeNET[28]. Using PsyGeNET we found that 26 of our DEGs have previously been linked to mental illness in the literature. The highest number of

associations (22/54 associations) were for depressive disorders, followed by associations for schizophrenia spectrum and other psychotic disorders (20/54; Fig. 3d). Using DisGeNET we found 15 genes associated with MDD related terms (hypergeometric test, p-value = 0.00029; Fig. 3e). Hypergeometric tests for overlap between DEGs in individual clusters and genes related to depression in DisGeNet revealed a specific enrichment in OPC2 DEGs ($p=5.7x10^{-4}$, Fig. 3e). Interestingly, we found that 67% of these genes were contributed by the OPC2 and Ex7 clusters (Fig. 3e). Complete results from PsyGeNET and DisGeNET are presented in Supplementary Table 33-35.

Functional implications of cell type specific DEGs

We used Gene Ontology and Reactome Pathway enrichment analysis to identify the relationship of our 96 DEGs to biological functions. There were strong enrichments of Gene Ontology terms for *neuron projection maintenance* (84-fold enrichment; FDR=0.011) and *negative regulation of long-term synaptic potentiation* (75-fold enrichment; FDR=0.012). Both of these terms are hierarchically related with the more general term *regulation of synaptic plasticity*, also enriched in the set of 96 genes (9-fold enrichment, FDR=0.012). Reactome Pathways enrichments included *Kinesins* (21.74-fold enrichment; FDR = $6.24x10^{-4}$), *HSP90 chaperone cycle for steroid hormone receptors* (15.79-fold enrichment; FDR = $3.4x10^{-2}$), and *Innate Immune System* (3.01-fold enrichment, FDR=$3.29x10^{-2}$). A full list of all enrichment analyses performed is provided in Supplementary Table 36-41.

The majority (excluding three*: AC133680.1, MEG3, FAM66C*) of the DEGs were protein-coding. We used STRING network analysis [29] to plot the interactions between these proteins coding DEGs. This enabled us to identify common pathways and systems, within which these proteins, contributed by different cell types, functionally interact. The overall connectivity between

proteins encoded by our DEGs was significantly higher than that expected for a random subset of genes (p-value = $3.64x10^{-4}$). While distinct genes were dysregulated in different clusters, common pathways and biological processes dysregulated across clusters included cytoskeletal function, immune system function, and SHR chaperone cycling (Fig. 4a), all of which have been previously implicated in MDD [9, 30].

Interestingly, certain genes were present in multiple pathways and processes, for example *HSP90AA1* (OPC2) links SHR chaperone cycling, immune system functioning and cytoskeletal function (Fig. 4b). Likewise, *KIF16B* from lower layer neurons (Ex7) and *KIF26B* and *KLC2* in two inhibitory cells types (In2 VIP and In3 SST respectively), belong to both the kinesin pathway and cytoskeletal function (Fig. 4c). Of note, *KAZN*, a gene previously associated with MDD[26], interacts with the *KIF16B* (Ex7), both of which represent some of the few upregulated genes in the dataset.

Weighted gene co-expression network analysis

In addition to directly measuring gene expression changes between groups, we performed weighted gene co-expression network analysis (WGCNA). To circumvent the challenges posed by the sparsity of snRNA-seq data, we performed WGCNA on the average gene expression profile for each subject across all cell types and included the percentage contribution of different cell types as a correlate. Our results indicated that 5 modules were significantly associated with MDD (Supplementary Table 42).

Four of the 5 modules were also strongly associated with Ex7, representing the highest cluster-phenotype overlap. We chose to focus on the largest module (blue), which included 2699 genes and significantly overlapped with our identified DEGs (Fig. 5a, 44%, p-value = $6.04x10^{-19}$,

hypergeometric test for overlap). To identify the most connected genes within the blue module, we performed a hub gene analysis resulting in 285 hub genes (Fig.5b, see Methods) and plotted the top 50, which included 10 DEGs (Fig. 5c). The top term for a Gene Ontology analysis of the hub gene list was "neurotransmitter secretion" (8.69-fold enrichment, FDR=7.21 $x10^{-3}$), suggesting a disruption of intercellular communication between neural cells. Furthermore, we found that 26 of the 41 DEGs that overlapped with the blue module were also hub genes (p-value = 4.95 x 10^{-31}, hypergeometric test for overlap).

Validation of gene expression changes

We preformed validation of our DEGs using fluorescence-assisted nuclei sorting (FANS) to separate broad cell types followed by high-throughput qPCR. As expected, given that the FANS fractions are much broader than the single cell clusters, with the 26 clusters combined into 4 sorted populations, levels of validation varied in part as a function of the relative representation of the cluster in the sorted fraction (Supplementary Fig. 10-11, Supplementary Tables 43-46). Figure 5 (d) highlights validated genes that overlap with the WGCNA results.

Intercommunication between lower layer excitatory neurons and oligodendrocyte precursor cells

Next, in order to better understand how cells are interacting, we applied a predictive tool to explore the relationship of ligands of one cluster to the receptors expressed in another cluster. We focused our analysis on Ex7 and OPC2, the two clusters showing the most DEGs, and with the greatest overlap of genes associated with phenotype from the literature and from our WGCNA. We found a total of 90 significantly changed ligand-receptor combinations between Ex7 and OPC2 after random permutations ($p<0.01$). Fifty-eight Ex7 ligand to OPC2 receptor (Fig.

6a left, Supplementary Table 47a) and 32 OPC2 ligand to Ex7 receptor interactions were altered between cases and controls (Fig. 6a right, Supplementary Table 47b). We found significant changes to FGF signalling originating from both cell types. Although these results are exploratory and need to be interpreted with caution, they are consistent with previous literature implicating the FGF system in MDD, and particularly, changes in FGF signalling in OPCs [31, 32] leading to depressive phenotypes, and provide an intriguing avenue for future experiments.

Based on the DEGs found in Ex7 and OPC2, we modeled the potential interaction indicating the class of protein and change in expression of the gene (Fig. 6b). To add support to the model we selected genes to further study with RNAScope® fluorescence *in situ* hybridization. Given the important change in FGF signalling we chose to investigate *FIBP* (FGF1 intercellular binding protein), *KAZN* a potential junction protein and HPS90AA1 a co-chaperone involved in stress hormone receptor cycling. We found *FIBP* was downregulated, as expected, in deep layer excitatory neurons (Fig. 6c, Unpaired t test, t_{217}=2.5, p=0.013, n= 95 nuclei for cases and controls) while *KAZN* was upregulated in OPCs (Fig. 6d, Unpaired t test, t_{188}=2.7, p=0.007, n=100 nuclei for controls, n=119 nuclei for cases) and *HSP90AA1* was downregulated, also in OPCs (Fig. 6e, Unpaired t test, t_{192}=2.0, p=0.026, n= 107 nuclei for controls, n= 87 nuclei for cases).

Discussion

Our examination of single-nucleus transcriptomes from the dlPFC in MDD revealed dysregulation of gene expression in almost 60% of the cell types identified, with a total of 96 differentially expressed genes. There were prominent gene expression changes in immature oligodendrocyte precursor cells (OPC2) and in deep layer excitatory neurons (Ex7), and a large percentage of their DEGs overlapped with genes previously implicated in MDD.

Given the complexity of psychiatric disorders such as MDD, disentangling the role of each cell type in the brain is important and requires single cell resolution. For example, the ability to distinguish glial subtypes – including multiple astrocytic, oligodendrocytic, and OPC clusters – enabled us to pinpoint changes specific to OPCs, but not oligodendrocytes, and changes selective to only one subset of astrocytic cells.

In recent years, the target cell types in MDD pathophysiology have expanded from excitatory neurons to include inhibitory interneurons[18] and non-neuronal cells[4-9]. Here we found 16 unique cell types showing evidence of differential gene expression in depression, including 4 non-neuronal clusters and 6 clusters of interneurons supporting the complex interplay between multiple cell types in MDD. Previous studies have shown that SST and PVALB interneurons are dysregulated in MDD patients[18], and here we report several DEGs in 3 interneuron clusters that are defined by the expression of these GABAergic markers (Inhib_3_SST, Inhib_6_SST, and Inhib_8_PVALB). Interestingly, a separate cluster of PVALB interneurons (Inhib_7_PVALB) did not show differential expression, which may indicate that not all PVALB interneurons are equally affected. However, we find differentially expressed genes in non-SST, non-PVALB interneuron clusters (Inhib_2_VIP, Inhib_1, and Inhib_5), which suggests that additional interneuron subtypes could have a role in depression, and should be examined in future research.

We found 10 different excitatory cell types which were annotated to specific cortical layers based on known markers. Ex10 represented a large cluster of superficial cortical layer cells, whereas there were numerous clusters representing different excitatory cell types from deeper cortical layers. The neuronal cluster with the most change was Ex7, a deep layer cluster characterized primarily by *DPP10* expression. *DPP10* encodes a dipeptidyl peptidase-related protein that regulates neuronal excitability and has previously been associated with a human-

specific, neuron-based regulatory network. Structural variants of this gene have been implicated in neuropsychiatric diseases, including autism, schizophrenia and bipolar disorder[33].

OPC2 also showed extensive gene expression changes between cases and controls. OPC2 was the youngest cell type in the OL pseudotime trajectory. The use of cellular deconvolution techniques indicated that OPC1 have some similarity to committed OPCs whereas OPC2 showed no such correspondence, supporting the idea of functional heterogeneity among OPCs [34]. Furthermore, compared to OPC1, OPC2 expressed higher levels of certain glutamate and sodium receptors, which are typically lost as the cells mature [34].

Evidence suggests that half of the OPCs ($NG2^{+}$) in the brain do not give rise to any other cell type[35], and exhibit synaptic contact with neurons [36]. As such, OPCs are now thought to be a distinct glial cell type implicated in brain plasticity through roles such as integration of synaptic activity[37] and mediation of long term potentiation[38]. Additionally, there is evidence directly implicating the loss of this cell type with emergence of depressive-like behaviour[31]. The data from this study support a role for OPCS in MDD independent from their role as precursor cells for oligodendrocytes.

STRING DB protein network analysis highlighted a number of links including connections between three differentially expressed genes encoding kinesin-related proteins: *KIF26B, KLC2* and *KIF16B*. *KIF16B* (increased in Ex7) is involved in recycling receptors including the fibroblast growth factor receptor (FGFR). Interestingly, *FIBP,* encoding acidic FGF1 intracellular-binding protein, was decreased in Ex7. FGFR transport relies, in part, on the interaction between kinesins and Rab GTPases [39]. Notably, we found *RAB11B* (encoding a Rab GTPase) and *KLC2* to be downregulated in In3. Taken together, these finding could point to a disruption of FGFR

recycling by kinesins and Rab GTPases, as well as disrupted modulation FGF intercellular signalling by *FIBP* in neurons in MDD.

Based on animal models and in cell culture, FGFs (specifically FGF2) and FGFRs seem to be affected by stress and the glucocorticoids[40]. The glucocorticoid receptor (GR) has consistently been implicated in MDD [41]. *HSP90AA1* (decreased in OPC2) and *FKBP4* (decreased in Ex7), along with its homolog *FKPB5*, encode cochaperones for the GR and regulate intracellular signalling functions of this receptor[30]. *HSP90AA1* codes for the stress inducible isoform HSP90α and interestingly, is known to be secreted in certain stress contexts [42]. These changes may point to a fundamental disruption in GR signaling in deep layer excitatory cells and OPCs, which could further interact with the above described changes in FGF signalling.

The genes related to chaperone mediated steroid hormone receptor cycling overlapped with genes involved in innate immune function. This is unsurprising given the role of glucocorticoids in modulating inflammation, one of the primary responses of the immune system. Both OPC2 and Ex7 were enriched for the common genes between these pathways. Finally both the FGF and GR system have implications in the plastic properties of excitatory neurons such as projection outgrowth and stability [43, 44].

Additionally, genes such as *PRNP* (the prion protein gene) and *KAZN* (a gene involved in desmosome assembly), were strongly altered in the OPC2 cluster and are associated with mediating synaptic plasticity and cellular communication[45, 46] . The absence of *Prnp* has been associated with an increased number of undifferentiated oligodendrocytes and the delayed expression of differentiation markers[47]. which is intriguing given the evidence implicating a lack of mature adult oligodendrocytes in animal models of depression and anxiety[48]. On the other hand, overexpression of kazrin in keratinocytes profoundly changed cell shape, reduced

filamentous actin, and impaired assembly of intercellular junctions [46]. Interestingly, decreased desmosome length has been described in *Prnp*$^{-/-}$ mice[49] suggesting an interplay between these proteins. Further, a SNP in *KAZN* showed one of the strongest associations in individuals with treatment resistant depression [26].

Based on the information we derived from various bioinformatics strategies we have proposed a putative model for the bidirectional interactions between lower layer excitatory neurons and immature oligodendrocytes. We used RNAScope® to validate some of the key transcriptional changes highlighted by the model. Though these results are interesting, functional follow up studies will be required to determine the role of molecules like FGF, HSP90α and Kazrin in the communication between these two cell types.

Our study is not without limitations. All individuals included in our study were male, so our results are not necessarily generalizable to women, particularly as previous studies have suggested that brain transcriptomic changes associated with MDD are different in females[50]. Nonetheless, this first screen provides important information that may help inform subsequent studies exploring both men and women with MDD. Technical limitations with droplet-based snRNA-seq of human brain have been previously described. We, like others[10, 11], found a much greater proportion of neurons compared to glial cells than would be expected based on histologically determined estimates, pointing to a potential limitation of the methodology for capturing non-neuronal cells. Although droplet-based snRNA-seq does not capture lowly expressed genes, nevertheless, we were able to perform differential gene expression for thousands of genes in precisely defined cell types.

Lastly, we believe the consistency across dissections was not sufficient for estimating cell type proportions. For example, even a small over-representation of one cortical layer versus another

during dissection, can give misleading results regarding the proportion of cell-types. Other groups have attempted to extract nuclei from cryo-sectioned samples to address these inconsistencies [10].

Our study has elucidated gene expression changes specific to numerous independent cell types in MDD. We have identified a potentially important link between OPCs and deep layer excitatory neurons, which implicates fundamental pathways including FGF signalling, glucocorticoid receptor regulation and synaptic plasticity in the brains of depressed individuals. The generalizability of these data will rely on independent validation in other MDD cohorts; nonetheless, this work provides an exciting start point for understanding the complex interplay of cells in the brain and a platform for future functional research to assess these potential interactions. Future single-cell studies of MDD should aim to relate cell types with symptomology and severity as has been done in recent papers [16, 17] .

Accession Codes

GEO accession number for snRNA-seq data: GSE144136.

Acknowledgements

GT holds a Canada Research Chair (Tier 1) and a NARSAD Distinguished Investigator Award. He is supported by grants from the Canadian Institute of Health Research (CIHR) (FDN148374 and EGM141899).

We acknowledge the expert help of the Douglas-Bell Canada Brain Bank staff (J. Prud'homme, M. Bouchouka and A. Baccichet), and and H. Djambazian at the MUGQIC. The work was also supported by CFI projects 32557 and 33408 (JR). The Douglas-Bell Canada Brain Bank is supported in part by funding from the Canada First Research Excellence Fund, awarded

to McGill University for the Healthy Brains for Healthy Lives project, and from the *Fonds de recherche du Québec - Santé* (FRQS) through the Quebec Network on Suicide, Mood Disorders and Related Disorders. The present study used the services of the Molecular and Cellular Microscopy Platform (MCMP) at the Douglas Institute.

Author Contributions

CN conceptualized, performed experiments and wrote the manuscript, MM performed experiments, bioinformatics and wrote manuscript. AT, VJ and MAD performed experiments and wrote the manuscript. MS, JFT, SJT and PP contributed to data analysis and reviewed the manuscript. NM contributed to tissue processing, data interpretation and manuscript preparation. JR provided technical single-cell expertise and experimental support, aided in manuscript preparation. GT provided general oversight of the study, including in experimental design, data interpretation and manuscript preparation.

Competing Interests statement

The authors declare no competing interests.

References

1. Organization, W.H. Depression and Other Common Mental Disorders: Global Health Estimates. 24 (Geneva, Switzerland, 2017).
2. Wray, N.R.*, et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668-681 (2018).
3. Jansen, R.*, et al.* Gene expression in major depressive disorder. *Molecular psychiatry* **21**, 339-347 (2016).
4. Sequeira, A.*, et al.* Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. *PLoS One* **4**, e6585 (2009).
5. Abdallah, C.G., Sanacora, G., Duman, R.S. & Krystal, J.H. The neurobiology of depression, ketamine and rapid-acting antidepressants: Is it glutamate inhibition or activation? *Pharmacol Ther* (2018).
6. Pantazatos, S.P.*, et al.* Whole-transcriptome brain expression and exon-usage profiling in major depression and suicide: evidence for altered glial, endothelial and ATPase activity. *Molecular psychiatry* **22**, 760 (2016).
7. Edgar, N. & Sibille, E. A putative functional role for oligodendrocytes in mood regulation. *Translational Psychiatry* **2**, e109 (2012).
8. Nagy, C.*, et al.* Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *Molecular psychiatry* **20**, 320-328 (2015).
9. Duman, R.S., Aghajanian, G.K., Sanacora, G. & Krystal, J.H. Synaptic plasticity and depression: new insights from stress and rapid-acting antidepressants. *Nat Med* **22**, 238-249 (2016).
10. Lake, B.B.*, et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80 (2018).
11. Habib, N.*, et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955-958 (2017).
12. Mancarci, B.O.*, et al.* Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data. *eneuro* **4** (2017).
13. Zheng, G.X.Y.*, et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 14049 (2017).
14. Lake, B.B.*, et al.* A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci Rep* **7**, 6031 (2017).
15. Renthal, W.*, et al.* Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nature Neuroscience* **21**, 1670-1679 (2018).
16. Mathys, H.*, et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
17. Velmeshev, D.*, et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).
18. Northoff, G. & Sibille, E. Why are cortical GABA neurons relevant to internal focus in depression? A cross-level model linking cellular, biochemical and neural network findings. *Molecular psychiatry* **19**, 966-977 (2014).
19. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
20. Sofroniew, M. & Vinters, H. Astrocytes: biology and pathology. *Acta neuropathologica* **119**, 7-35 (2010).

21. Anderson, M.A., Ao, Y. & Sofroniew, M.V. Heterogeneity of reactive astrocytes. *Neurosci Lett* **565**, 23-29 (2014).
22. Trapnell, C*., et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).
23. Thomas, P.D*., et al.* Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res* **34**, W645-650 (2006).
24. Butts, B.D., Houde, C. & Mehmet, H. Maturation-dependent sensitivity of oligodendrocyte lineage cells to apoptosis: implications for normal development and disease. *Cell Death Differ* **15**, 1178-1186 (2008).
25. Jakel, S*., et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543-547 (2019).
26. Li, Q.S., Tian, C., Seabrook, G.R., Drevets, W.C. & Narayan, V.A. Analysis of 23andMe antidepressant efficacy survey data: implication of circadian rhythm and neuroplasticity in bupropion response. *Transl Psychiatry* **6**, e889 (2016).
27. Gutierrez-Sacristan, A*., et al.* PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics* **31**, 3075-3077 (2015).
28. Pinero, J*., et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).
29. Szklarczyk, D*., et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).
30. Wochnik, G.M*., et al.* FK506-binding proteins 51 and 52 differentially regulate dynein interaction and nuclear translocation of the glucocorticoid receptor in mammalian cells. *J Biol Chem* **280**, 4609-4616 (2005).
31. Birey, F*., et al.* Genetic and Stress-Induced Loss of NG2 Glia Triggers Emergence of Depressive-like Behaviors through Reduced Secretion of FGF2. *Neuron* **88**, 941-956 (2015).
32. Mason, J.L. & Goldman, J.E. A2B5+ and O4+ Cycling progenitors in the adult forebrain white matter respond differentially to PDGF-AA, FGF-2, and IGF-1. *Mol Cell Neurosci* **20**, 30-42 (2002).
33. Shulha, H.P*., et al.* Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS Biol* **10**, e1001427 (2012).
34. Spitzer, S.O*., et al.* Oligodendrocyte Progenitor Cells Become Regionally Diverse and Heterogeneous with Age. *Neuron* **101**, 459-471 e455 (2019).
35. Psachoulia, K., Jamen, F., Young, K.M. & Richardson, W.D. Cell cycle dynamics of NG2 cells in the postnatal and ageing brain. *Neuron Glia Biol* **5**, 57-67 (2009).
36. Bergles, D.E., Jabs, R. & Steinhauser, C. Neuron-glia synapses in the brain. *Brain Res Rev* **63**, 130-137 (2010).
37. Birey, F., Kokkosis, A.G. & Aguirre, A. Oligodendroglia-lineage cells in brain plasticity, homeostasis and psychiatric disorders. *Current Opinion in Neurobiology* **47**, 93-103 (2017).
38. Ge, W.P*., et al.* Long-term potentiation of neuron-glia synapses mediated by Ca2+-permeable AMPA receptors. *Science* **312**, 1533-1537 (2006).
39. Ueno, H., Huang, X., Tanaka, Y. & Hirokawa, N. KIF16B/Rab14 molecular motor complex is critical for early embryonic development by transporting FGF receptor. *Dev Cell* **20**, 60-71 (2011).
40. Turner, C.A., Watson, S.J. & Akil, H. The fibroblast growth factor family: neuromodulation of affective behavior. *Neuron* **76**, 160-174 (2012).

41. Turecki, G. & Meaney, M.J. Effects of the Social Environment and Stress on Glucocorticoid Receptor Gene Methylation: A Systematic Review. *Biol Psychiatry* **79**, 87-96 (2016).
42. Zuehlke, A.D., Beebe, K., Neckers, L. & Prince, T. Regulation and function of the human HSP90AA1 gene. *Gene* **570**, 8-16 (2015).
43. Huang, J.Y., Lynn Miskus, M. & Lu, H.C. FGF-FGFR Mediates the Activity-Dependent Dendritogenesis of Layer IV Neurons during Barrel Formation. *J Neurosci* **37**, 12094-12105 (2017).
44. Pittenger, C. & Duman, R.S. Stress, depression, and neuroplasticity: a convergence of mechanisms. *Neuropsychopharmacology* **33**, 88-109 (2008).
45. Caiati, M.D*., et al.* PrPC controls via protein kinase A the direction of synaptic plasticity in the immature hippocampus. *J Neurosci* **33**, 2973-2983 (2013).
46. Sevilla, L.M., Nachat, R., Groot, K.R. & Watt, F.M. Kazrin regulates keratinocyte cytoskeletal networks, intercellular junctions and differentiation. *J Cell Sci* **121**, 3561-3569 (2008).
47. Bribian, A*., et al.* Role of the cellular prion protein in oligodendrocyte precursor cell proliferation and differentiation in the developing and adult mouse CNS. *PLoS One* **7**, e33872 (2012).
48. Liu, J*., et al.* Impaired adult myelination in the prefrontal cortex of socially isolated mice. *Nat Neurosci* **15**, 1621-1623 (2012).
49. Morel, E*., et al.* The cellular prion protein PrP(c) is involved in the proliferation of epithelial cells and in the distribution of junction-associated proteins. *PLoS One* **3**, e3000 (2008).
50. Labonte, B*., et al.* Sex-specific transcriptional signatures in human depression. *Nat Med* **23**, 1102-1111 (2017).

Figure Legends

Figure 1: Experimental Flow. Schematic representation of experimental procedures. Nuclei were extracted from Brodmann area 9 (BA9) in the dlPFC of 17 cases and 17 controls, single nuclei were captured in droplets for RNA-seq. Unsupervised clustering and cell type annotation were followed by differential expression analysis between cases and controls within each cluster. Bioinformatic analyses were performed to link the changes to the phenotype. Two validation approaches: FANS-high-throughput qPCR and FISH, were applied for validating differential expression results.

Figure 2: Identification of cell types a) TSNE plot depicting the ~73,000 cells in 26 clusters identified after strict quality control of initial clusters.**b)** Cell type annotation was performed based on expression of well-established marker genes. (Left) Dendrogram representing relationship between identified cell type clusters based on gene expression. (Middle) DotPlot depicting expression of known marker genes in the 26 clusters of interest. Marker genes are colour coded according to the cell type in which they should be detected. The size of the dots represents the proportion of cells expressing the gene whereas the colour intensity represents the average expression level. (Right) Columns listing the number of cells per group and the bar plot depicting the mean number of UMIs per cell in each cluster. **c)** Cortical layer specific markers varied in expression within the excitatory neuronal clusters. The violin plots depict the expression per cluster of layer specific marker genes going from the more superficial layers (I/II) on the left to the deeper layers (V/VI) on the right. **d)** Known classes of inhibitory neurons are identifiable based on the expression pattern of peptide genes (*VIP, SST, CCK*) and calcium binding protein genes (*PVALB*). **e)** (Left, violin plots) Cells belonging to the oligodendrocyte lineage expressed the expected markers**.** (Top) The oligodendrocyte lineage cells from 5 clusters were analysed to produce a pseudotime trajectory to gauge their developmental stages. . (Right)

The location of these clusters along the trajectory was consistent with deconvolution (Jäkel et al., 2019). The numbers represent the percentage contribution of each of the previously published cluster signatures to the corresponding clusters in our dataset. For violin plots in figures 2c-e values extend from minimum to maximum, the median value is indicated by a dot and the n-value per cluster corresponds to the total "No. of cells"for cases and controls combined listed in 2b. Nuclei were derived from 34 subjects.

Figure 3: Differentially expressed genes. a) For each cluster the percentage change in expression between cases and controls of all detected genes are plotted with decreased expression to the bottom of the midline and increased expression to the top. Ninety-six significantly changed genes (16 were up-and 80 down-regulated) are marked in colour, based on their corrected FDRs as shown in the legend. The numbers of nuclei from cases and controls per cluster (n) are available in Supplementary Tables 8-31. p-values were obtained using a mixed linear model (see Methods). Nuclei were derived from 34 subjects. Sixteen out of the 26 clusters contained significantly differentially expressed genes. (Insert) Stacked bar-graph shows contribution of different cell type clusters to differentially expressed genes. **b)** Number of clusters in each broad category showing up and downregulated genes in MDD cases. **c)** The scatter plots represent the number of DEGs and the average percentage change in expression for each cluster. The cluster size is depicted by the size of the circle. Upper graph depicts upregulated genes, lower graph depicts downregulated genes. OPC2 and Ex7 show the highest level of both up and down regulated genes. **d-e)** The number of genes with known relationship to psychiatric phenotypes using available databases PsyGeNET and DisGeNET. **d)** 26 of the 96 dysregulated genes were found in PsyGeNET

and showed an enrichment for MDD (Total, all the genes which overlap database for a given disorder; 100% association, the genes positively associated with the disease; 100% no association, the genes negatively associated with the disease; both, mixed findings (positive and negative) for a given gene related to the disease. **e)** (Left) 15 genes were found to be associated with depression related terms in DisGeNET. (Right) The percentage of genes per cluster associated with MDD from DisGeNET, along with cluster specific enrichment of DisGeNET MDD associated genes. For hypergeometric tests, the number of depression-associated genes in DisGeNET was 1199 and the number of unique genes in DisGeNET was 17545 for all tests. The number of DEGs in DisGeNET (k) and the number of depression-associated DEGs (x) are listed: All clusters: k=85, x=15; OPC2: k=24, x=7; Ex7: k=19, x=3; Endo: k=2, x=1; Astro3: k=6, x=1; Ex3: k=2, x=1; In2: k=11, x=1; In5: k=2, x=1.

Figure 4: Differential expression and biological associations. a) String DB network for all DEGs with nodes corresponding to a set of biological processes and pathways highlighted (legend on right). **b)** Subset of genes shared between the immune function related terms and the steroid hormone receptor cycling pathway. **c)** Subset of genes involved in cytoskeletal function and kinesin activity. Colour strips beneath networks give a proportional representation of the contributing clusters.

Figure 5: Weighted gene co-expression network analysis. a) Venn diagram of overlap between blue module genes and DEGs (hypergeometric test, p-value = 6.037692e-19). **b)** Venn diagram for overlap between blue module hub genes and DEGs (hypergeometric test, p-value = 4.954172e-31). **c)** Visualization of the top 50 hub genes assessed for the blue module. DEG nodes and all edges connected to them are colored teal. **d)** Boxplots represent expression levels of DEGs validated with high-throughput qPCR in FAN sorted populations which were also hub

genes in the blue module. Mann-Whitney U tests (two-sided) were performed for *PRAF2* as the values were not normally distributed based on the Shapiro Wilk's test for normality. All other genes were tested with unpaired two-sided t-tests as their values were normally distributed. P-values: * < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001. Whiskers on box plot represent maximum and minimum values. Box extends from the 25th percentile to 75th percentile, the center line represents the median, and dots represent all values in the dataset. *ATP6V0B*: n=15 cases, 11 controls, t=3.10, df=12.62, p-value=0.0087; *CKB*: n=9 cases, 7 controls, t= 2.48, df=16.85 p-value= 0.023 ; *PRAF2*: n=14 cases, 10 controls, U= 8, p-value=6.8 x 10^{-5}; *TKT*: n= 16 cases, 14 controls, t= 2.25, df=19.83, p-value=0.036; *PLD3*: n=15 cases, 14 controls, t= 3.06, df= 15.83, p-value=0.0075; *OTUB1*: n=16 cases, 14 controls, t= 2.39, df=20.92, p-value=0.026; *ACTB*: n=14 cases, 15 controls, t=3.14, df= 19.98, p-value=0.0052; *HNRNPK*: n= 14 cases, 13 controls, t=2.41, df=16.07, p-value=0.028.

Figure 6: Contributions of OPC2 and Ex7. a) CCInx receptor ligand based cell-cell interaction network analysis for communication between Ex7 and OPC2. Given the large number of connections (Supplementary Tables 47a,b), a subset are shown. **b)** Our data points to a change in the communication between deep layer excitatory neurons (Ex7) and immature OPCs (OPC2). Altered FGF bidirectional signalling was identified via CCInx. We propose that immature OPCs have a very important role in regulating plastic properties of deep layer excitatory cells, such as neuron projection outgrowth and maintenance. Lines between cell types are labeled with secreted or junction proteins found to be dysregulated in the given cell type for example *HPS90AA1* codes for the stress inducible isoform HSP90α, known to be secreted in certain contexts, *KAZN* is an upregulated junction protein *i*n OPCs and *ATP6V0B* could represent altered ATP signaling. Arrows beside gene names indicate up or downregulation. Beside each cell type are the genes in given functional categories and their direction of change in the disease state. **c)**

Decreased expression of the gene encoding FGF1 intercellular Binding Protein (*FIBP*) was validated in deep layer neurons using RNAScope®. *SLC17A7* (encoding VGLUT) was used as a marker for excitatory cells and *RXFP1* was used to identify deep layer neurons. *SLC17A7*$^{+}$*, RXFP1*$^{+}$ cells were imaged and *FIPB* expression was counted (Cases: n=119 nuclei, controls: n=100 nuclei, unpaired two-sided t-test, t = 2.49, df= 217, p = 0.013). **d)** Increased *KAZN* (cases: n=95 nuclei, controls: n=95 nuclei, unpaired two-sided t-test, t = -2.69, df= 188, p = 0.008) and **e)** decreased *HSP90AA1* (cases: n = 87 nuclei, controls: n = 107 nuclei, unpaired two-sided t-test, t = 2.23, df= 186, p= 0.027 expression were validated in OPCs using *PDGFRA* as a marker for oligodendrocyte precursor cells. Whiskers on box plot represent the 5th and 95th percentile. Box extends from the 25th percentile to 75th percentile and the center line represents the median. Dots represent points beyond the 5th or 95th percentile. Scale bar represents 5µm.

Tables

Table 1: Sample information

	Controls (n=17)	Cases (n=17)	p value
Age (years)	38.71 ± 4.32	41.06 ± 4.66	p=0.714
Gender	17M	17M	-
PMI (hrs)	34.01 ± 4.94	41.69 ± 4.76	ǂp=0.190
pH	6.49 ± 0.06	6.60 ± 0.07	p=0.212
Storage Time (years)	14.71± 1.44	12.47± 1.46	ǂp=0.543
Cause of death	Accident (6), Natural (11)	Suicide (17)	
Substance dependence	None	None	
Comorbid diagnoses	None	None	
Toxicology	EtOH (2), Cannabinoids (1),	EtOH (6), BZ (1), AD (2), Cannabinoids (1), Cocaine (1),	
Antidepressant Treatment	None	3	

Mean ± SEM
ǂMann Whitney test
NA – not applicable, EtOH – ethanol, BZ – benzodiazepines, AD – antidepressants, AC – anticonvulsants

Materials and Methods

Subjects: Postmortem brain samples

This study was approved by the Douglas Hospital Research Ethics Board, and written informed consent from next-of-kin was obtained for each subject. Postmortem brain samples were provided by the Douglas-Bell Canada Brain Bank (www.douglasbrainbank.ca). Frozen grey matter samples were dissected from Brodmann Area 9 (dlPFC). Brains were dissected by trained neuroanatomists and stored at -80 °C. For each individual, the cause of death was determined by the Quebec Coroner's office, and psychological autopsies were performed by proxy-based interviews, as described previously[51]. Cases met criteria for MDD and died by suicide whereas controls were individuals who died suddenly and did not have evidence of any axis I disorders (Table 1). Post mortem interval (PMI) represents the delay between a subject's death and collection and processing of the brain. To assess RNA quality, we measured the RIN obtained for our samples using tissue homogenates. An unpaired, two-tailed, Student's t-test revealed no significant difference (p=0.15) in RIN between cases (mean RIN of 6.74) and controls (mean RIN of 6.16). 17 cases and 17 controls were included in the snRNA-seq experiment and the full cohort of subjects (except 25) was used for follow-up validation of DEGs by FANS and high throughput qPCR. RNAScope experiments were performed on representative subsets of samples using 5 cases and 5 matched controls. Detailed information on experimental design and reagents can also be found in the Life Sciences Reporting Summary.

Nuclei isolation and capture

50 mg of frozen tissue was dounced in 3 mL of lysis buffer, 10 times with a loose pestle and an additional 5 times with the tight pestle. The lysis buffer contained 10 mM Tris (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, and 0.05% (v/v) NP-40 detergent. The sample was left to lyse in a total of 5

mL of buffer for 5 min, after which 5 mL of wash buffer was added and swirled. The sample was passed through a 30 µm cell strainer and spun for 5 min at 500 *g*. This step was repeated for a total of two filtering steps. After pelleting, the nuclei are resuspended in 5-10 mL of wash buffer by pipetting up and down 8-10 times. After 3 washes, the nuclei were resuspended in 1 mL of wash buffer and mixed with 25 % Optiprep™ and layered on a 29 % Optiprep™ cushion and spun for 30 min at 10,000 *g*. Nuclei were resuspended in wash buffer to achieve a concentration of ~$1x10^6$ nuclei/mL. Representative images of extracted nuclei are presented in Supplementary Fig. 12.

We used the 10x Genomics® Chromium™ controller for single cell gene expression to isolate single nuclei for downstream bulk RNA library preparation. We strictly followed the protocol as outlined by the user guide (***CG00052_SingleCell3_ReagentKitv2UserGuide_RevE.pdf***), with the exception of loading concentration, which we increase by 30% as we assessed the capture of nuclei to be slightly less efficient than cell encapsulation. We aimed to capture ~3000 nuclei per sample. So, for example, if our sample concentration was 390 nuclei/µL (~ 400 nuclei/ µL) according to page 10 of Protocol Step 1 we are required to load 13.1 µL of the stock to capture 3000 cells. But instead, we would recalculate our stock concentration to be 70% of 390 = 273 nuclei/ µL and load 17.4 µL (the recommended amount for 300 nuclei/ µL) instead. This system only allows for a maximum of 8 samples per capture run. As such, we required multiple batches to collect the individual nuclei for all 34 samples (6 batches). Samples 24 and 25 performed poorly, we therefore, carried out the capture on two separate chips and sequenced twice combining the data from both runs for the final analysis.

Sequence Alignment and UMI Counting

A pre-mRNA transcriptome was built using the cellranger mkref (Cellranger version 2.0.1) command and default parameters starting with the refdata-cellranger-GRCh38-1.2.0 transcriptome and as per the instructions provided on the 10X Genomics website. Reads were demultiplexed by sample index using the cellranger mkfastq command (Cellranger v2.1.0). Fastq files were aligned to the custom transcriptome, cell barcodes were demultiplexed, and UMIs corresponding to genes were counted using the cellranger count command and default parameters.

Data Transformation for Secondary Analysis

The unfiltered gene barcode matrices for each sample were loaded into R using the Read10X function in the Seurat R package (version 2.2.0, 2.3.0)[19]. Cell names were modified such that the subject name, batch, and biological condition were added to them. Seurat objects were created corresponding to each sample using the CreateSeuratObject function with the imported unfiltered gene-barcode matrices provided as the raw data. Individual Seurat objects for each sample were combined into one object using the MergeSeurat function sequentially. No filtering or normalization was performed up to this step. Since this is a single nucleus dataset, all mitochondrial genes that are transcribed from the mitochondrial genome were removed, along with genes not detected in any cell.

Barcode and Gene Filtering

Based on the distribution of nGene (total number of genes detected in each cell) for the total dataset (assessed by summary and hist R[52] functions), barcodes that were associated with less than 110 detected genes were removed. Based on the distribution of nUMI (total numbers of UMIs detected in each cell), the top 0.5 % of barcodes were also excluded as most likely being

multiplets rather than single nuclei, as there was a very sharp increase of nUMI from 16,393 at the 99.5th percentile to 102,583 at the maximum.

Next, the distribution of nUMI for the remaining barcodes was fit with three normal distributions using the normalmixEM function from the mixtools[53] package (Supplementary Fig. 1c). The rationale was that, the filtered barcodes contain a population of low quality "noise" barcodes that have a very low nUMI on average, a population of non-neuronal cells that have an intermediate nUMI and a population of neuronal cells that have a high nUMI. Based on the fitting of the normal distributions, only the barcodes with a high probability (> 0.95) of belonging to either the putative "non-neuronal" or putative "neuronal" distributions, and a low probability (<0.05) of belonging to the "noise" distribution were retained for further analysis (Supplementary Fig. 1c-d). 78,886 cells and 30,062 genes were retained.

Our custom filtering (Supplementary Fig. 1a-e, Supplementary Table 4) helped to increase the number of glial cells recovered. With an initial subset of 20 subjects, applying our custom filtering increased the total number of cells 1.8–fold but increased the number of non-neuronal cells by almost 6-fold (data not shown). After custom filtering the minimum numbers of genes and UMIs per nucleus were 254 and 340 respectively.

Once nuclei were filtered, the percentages of mitochondrial reads associated with the retained barcodes were calculated although for quality control purposes those reads were not used during the filtering or downstream analysis (Supplementary Fig. 1f). Although the percentage of reads mapping to mitochondrially expressed genes is a more pertinent quality control parameter for single-cell rather than single nucleus approaches, contaminating mitochondrial reads often present a problem in single-nucleus protocols (pers. comm., Lake, B.B.). However, our optimized approach was able to minimize this technical issue.

Data Processing and Dimensionality Reduction

The UMI counts were normalized to 10,000 counts per cell and converted to log scale (Seurat function NormalizeData). The batch, condition, and subject information was added as metadata to the final Seurat object; nUMI and batch were regressed out using the ScaleData function. The Seurat FindVariableGenes function was used with default selections and cut-offs as follows: x.low.cutoff = 0.003, x.high.cutoff = 2, y.cutoff = 1. This resulted in a list of 2135 highly variable genes, which excludes lowly expressed genes (below 25^{th} percentile), very highly expressed genes, and selects only the top 10 % of genes in terms of the scaled dispersion. These highly variable genes were used to calculate 100 principal components. Based on the PC elbow plot of the standard deviation of the PCs (Supplementary Fig. 2a), the first 50 PCs were retained for use in downstream analysis.

Clustering by Gene Expression

The FindClusters function was applied with a resolution of 2.5 and produced 44 initial clusters. The goal of clustering is to sort nuclei by cell type so that all remaining gene expression variation within clusters is not related to cell differentiation processes. Prior to the advent of single nuclei expression profiling, cell types were identified by observing differences in cell morphology, behaviour, and anatomic location. It is fairly straight-forward to sort single nuclei expression profiles into known cell types according to the expression levels of marker genes that differentiate between these cell types. However, it is very unlikely that all cell types have been identified so we must rely on nuclei clustering to uncover as-yet unknown cell types. Unfortunately, the number of clusters obtained from the clustering algorithm is somewhat arbitrary because clustering depends on the settings of several parameters, and there is no consensus on how they should be set. Although clusters obtained using reasonable default

settings usually correspond to known biological cell types, some clusters may appear to potentially identify entirely new cell types or splinter existing cell types into multiple subtypes. Deciding if the clusters really do identify new cell types can be difficult or may even be impossible from available data.

To address this issue, we used tools in the Seurat package to sequentially combine any clusters that were not sufficiently distinct from each other. In particular, after performing initial hierarchical clustering of the graph-based clusters (BuildClusterTree), we assessed the nodes of the dendrogram using a random forest classifier (AssessNodes) and then merged together any nodes which were in the bottom 25 % of the dendrogram (using the branching.times function from the ape R package[54]) and had an out-of-bag-error of more than 5 %. We then repeated this clustering and merging process for the nuclei within each terminal node until none of the remaining nodes fulfilled our cut-off criteria (Supplementary Fig. 2b). The resulting set of 30 clusters were then characterized in terms of known markers genes of all major, well-defined brain cell types (Supplementary Fig. 2c-d). For refining identification of excitatory neuron types, we combined and re-clustered a set of excitatory clusters with highly correlated gene expression profiles ($R > 0.95$) (Supplementary Fig. 13a-c) using similar parameters for clustering as the whole dataset. This included 7 clusters of ~40,000 cells. Reclustering yielded 33 final clusters for downstream analysis. Finally, the clusters were manually curated to eliminate potential biases; for example, clusters were removed if mainly one sample contributed to the cells contained within the cluster (Supplementary Tables 48-51, Supplementary Fig. 14a-e).

Cluster Annotation

Genes used as markers for major cell-types and layer-specificity are listed below. Inhibitory neuron subtypes were annotated based on expression of canonical inhibitory interneuron

markers *SST, PVALB,* and *VIP* where possible. Excitatory neuron subtypes were annotated with some level of layer specificity based on expression of layer specific markers[11, 55, 56]. We also characterised clusters in terms of all genes differentially expressed between clusters (FindAllMarkers function, bimodal test, logfc.threshold of log(2), other parameters set to default) (Supplementary Table 6).

Major cell-type markers (Supplementary Fig. 3a-p)

Macrophage/ Microglia: *SPI1, MRC1, TMEM119, CX3CR1*; **Endothelial:** *CLDN5, VTN;* **Astrocytes:** *GLUL, SOX9, AQP4, GJA1, NDRG2, GFAP, ALDH1A1, ALDH1L1, VIM;* **OPCs***: PTGDS, PDGFRA, PCDH15, OLIG2, OLIG1;* **Oligodendrocytes:** *PLP1, MAG,MOG, MOBP, MBP;* **Excitatory neurons:** *SATB2, SLC17A7, SLC17A6;* **Inhibitory neurons:** *GAD1,GAD2, SLC32A1;* **Neurons:** *SNAP25,STMN2, RBFOX3.*

Layer-specific markers:

L2: *GLRA3;* **L2-3:** *LAMP5, CARTPT*; **L2-4:** *CUX2, THSD7A*; **L2-6:** *RASGRF2, PVRL3*; **L3-4**: *PRSS12;* **L4-5**: *RORB*; **L4-6:** *GRIK4*; **L5**: *KCNK2, SULF2, PCP4, HTR2C, FEZF2*; **L5-6**: *TOX, ETV1, RPRM, RXFP1, FOXP2;* **L6:** *SYT6, OPRK1, NR4A2, SYNPR, TLE, NTNG2, ADRA2A*

Pseudotime trajectory using Monocle

For oligodendrocyte developmental trajectory assessment, the data for cells belonging to the five clusters in the oligodendrocyte lineage (Oligos_1, Oligos_2, Oligos_3, OPCs_1, OPCs_2) were used to create a separate Seurat object using the SubsetData function. The most variable genes for these clusters alone were identified using the FindVariableGenes function and the following parameters: x.low.cutoff = 0.003, x.high.cutoff = 3, y.cutoff = 1 (giving a total of 895). The Seurat object was imported into a CDS (CellDataSet) object using the Monocle[22] function importCDS.

Estimation of size factors and dispersions was performed (using the estimateSizeFactors and estimateDispersions Monocle functions) on the CDS object using default parameters. Dimensionality reduction was then performed using reduceDimension, with reduction_method set to DDRTree. The 895 variable genes identified as above were used for ordering the cells into a trajectory with the orderCells function. The pseudotime trajectory was then plotted with plot_cell_trajectory (Fig. 2e), and the change in expression of genes known to be involved in oligodendrocyte development were plotted using plot_genes_in_pseudotime (Supplementary Fig. 6b-i). differentialGeneTest was applied separately to oligodendrocyte lineage cells from control subjects and MDD cases with fullModelFormulaStr = "~sm.ns(Pseudotime)". This allows us to model the expression of each gene as a function of pseudotime. All genes detected in at least one cell in the respective group were compared and their changes across pseudotime were assessed. A q-value cut-off of < 0.01 was used to identify genes associated with pseudotime. The overlapping and non-overlapping genes were identified by comparing the lists obtained for the two groups (Supplementary Fig. 6a).

Purification of Clusters for Differential Expression

Our doublet removal approach comprised of calculating a median gene expression profile for all our clusters, calculating the correlation of the gene expression of each cell, with the median profile of its cluster (considering only the top 865 genes whose median expression was highly variable, that is had a variance of > 0.25 across the different cluster) and selecting cells with high correlation. This was done by fitting bimodal normal distributions to the total distribution of correlations in the cluster to identify low and high correlation peaks. Cells were retained only if they had a low probability of falling in the low correlation peak ($p < 0.25$) and a high probability ($p > 0.75$) of falling in the high correlation peaks (Supplementary Fig. 7).

Differential Gene Expression Analysis

Differential expression analysis between the cases and controls was performed using linear mixed models implemented in the lme4[57] and lmerTest[58] R packages. Mixed models were necessary in order to account for dependencies between nuclei obtained from the same subject. Biological condition and number of UMIs were included in models as fixed effects and the subject and batch as random effects. The inclusion of subject as a random effect should account for subject specific effects such as age and PMI as well as technical effects of capture and library preparation which was performed separately for each subject. A false discovery rate (FDR) of 0.1 was used to detect differentially expressed genes within each cell type.

Weighted Gene Co-expression Network Analysis (WGCNA)

Average cell-expression for each sample across every cluster was calculated. These average counts were converted to log + 1 counts to reduce dispersion. WGCNA analysis was carried out in R with the WGCNA package (version 1.68) by Langfelder and Horvath. Genes with insufficient variance were excluded as well as outlier samples. After some tests, a soft-thresholding power of 7 and a minimum module size of 30 genes were selected for the gene network construction. Resulting modules were correlated with the phenotype information (MDD vs Control), as well as each sample's respective composition of each of the 26 single-cell type clusters they're composed of.

We performed hub gene analysis on the blue module, which was the largest module (2699 genes) which was correlated to phenotype. Potential hub genes were identified in the module of interest my selecting genes with a module membership larger than 0.80 and a gene significance larger than 0.20 with a p-value of less than 0.05. The top 50 potential hub genes were extracted

alongside any weighted interaction of more than 0.2. The resulting network was visualized in Cytoscape (3.7.1).

Fluorescence-assisted nuclei sorting (FANS)

Nuclear suspensions were prepared from 80-100 mg of post-mortem brain tissue from BA9 as described previously [59] with the following modifications: homogenized tissue was centrifuged on the sucrose layer at 800g for 20 minutes at 4^{o}C, followed by another centrifugation in nuclei extraction buffer. Resuspended nuclei were stained with the following primary antibodies in 600 µL of blocking buffer: mouse anti-CUTL2-PerCP conjugated (1:100, Novus catalog no. H00023316-M03, clone 2H8, conjugated to PerCP using the Novus Lightning Link Labeling kit, catalog no. 718-0010), goat anti-SOX10 (1:100, R&D Systems catalog no. AF2864), mouse anti-NeuN-A700 (1:300, Novus catalog no. NBP1-92693AF700, clone- 1B7) by incubating at room temperature, away from light, with rotation for 2 hours. Secondary antibody (donkey anti-goat Alexa Fluor 488, 1:1000, JacksonImmuno 705-545-147) was added and incubated for 1 hour at room temperature with rotation. All antibodies were purchased from Cedarlane. Nuclei were washed with PBS and the DNA was stained by Hoechst 33342 (Invitrogen, H1399).

FACSAria Fusion (BD Biosciences, San Jose, CA) was used for sorting of four populations – SOX10 positive, SOX10 negative, CUTL2 positive and CUTL2 negative. Gating strategy for the sorts is shown in (Supplementary Fig. 11) and was as follows. Doublet discrimination was achieved by gating of Hoechst 33342 stained singlets in FSC-A vs Hoechst-A plot using 350 nm UV laser and 450/50 filter. Subsequent SOX10 positive, SOX10 negative and NeuN positive populations were gated in Alexa Fluor 700-A vs Alexa Fluor 488-A plot utilizing red 640 nm laser in combination with 730/45 filter and blue 488 laser in combination with 530/30 filter, respectively. CUTL2 positive and negative populations, the derivatives of NeuN positive gate, were defined in Alexa

488-A vs PerCP-A (blue 488 laser, 695/40 filter) plot with interval gates. CUTL2 positive population was identified as 30-40% of NeuN positive population with highest CUTL2-PerCP fluorescence. For gating of CUTL2 negative population the SOX10 negative and SOX10 positive populations were displayed in Alexa 488-A vs PerCP-A plot and the CULT2 negative population was gated within PerCP intensities of SOX10 populations. CUTL2 negative population comprised near 10% of NeuN positive population.

Validation information for antibodies is as follows: Novus H00023316-M03- validated in Western blot and ELISA, used in one publication in human brain tissue (PMID: 29126813); R&D Systems AF2864- validated in Western blot against human SOX10 protein, ELISAs, immunocytochemistry, 19 citations; Novus NBP1-92693AF700- validated in immunocytochemistry, immunohistochemistry, Western blot, one publication for flow cytometry in human brain tissue (PMID: 28750583).

High-throughput qPCR

RNA was extracted from FANS sorted nuclei population using the Norgen RNA/DNA Purification Kit (Cat. 48700). cDNA was synthesized using a modified SMART-seq procedure as described previously [60]. The Fluidigm Biomark system was used for performing high-throughput qPCR as per manufacturer protocol as previously described[61]. Fludigim Delta Gene™ primer designs were used for the 93 targets (all differentially expressed transcripts excluding *AC133680.1*) and 3 endogenous controls (*GAPDH, POLR2A, UBC*).

Cell-cell interaction measurement

To assess cell-cell communication, we calculated predicted ligand-receptor interactions between Ex7 and OPC2 using CCInx[62] (https://github.com/BaderLab/CCInx), in which the connection

between each ligand and receptor is quantified as an edge weight. We chose a gene expression threshold of 2.75 and above to limit our research to relatively highly expressed ligands and receptors and for ease of visualization. To test if the edge weights were significantly different between cases and controls, we randomly permuted our subjects into two groups 100 times and formed normal distributions of the edge weight differences between groups for each ligand-receptor pair. We then calculated a p-value for the case-control edge weight difference for each ligand-receptor pair based on its position in the distribution. Edge weight difference p-values <0.01 were considered significant. A sample script used for assessing the significance of edges has been provided.

Cell deconvolution for all clusters

Expression data from (dbGaP:phs000424.v8.p1)[11] was used as reference signatures for annotated cell types. UMI counts for each cell were converted to transcripts per million (TPMs) in order to account for the varying sequencing depth of each cell and sample. Average expression levels were calculated for each cell type-specific cluster defined in the paper.

Cluster-specific gene expression profiles were obtained by summing the UMI values of all 24301 genes common to our dataset and the reference for each nucleus in each cluster and converting the sums to TPMs. R package, DeconRNASeq v1.18.0[63] was used to deconvolute these cluster-specific profiles. Using the data from[11]as reference, we were able to estimate the cell type composition of our clusters.

Cell deconvolution for oligodendrocyte lineage

Average expression from every control samples from the Jäkel et al. dataset were calculated and used as cell signatures for the deconvolution of our oligodendrocytic clusters (average cell

expression of every cell in the cluster considered as bulk) using the R package DeconRNASeq (v 1.26.0).

RNA-Scope Fluorescent *In Situ* Hybridization

Frozen BA9 blocks were cut serially with a cryostat (10µm thickness) on superfrost charged slides and kept at -80^{o}C until further processed. *In situ* hybridization was performed using Advanced Cell Diagnostics RNAscope® probes and reagents according to the manufacturer instructions in 5 matched subjects per group. Briefly, sections were first fixed in chilled 10% neutral buffered formalin for 15 mins at 4^{o}C, dehydrated by increasing gradient of ethanol bathes and left to air dry for 5 minutes. Endogenous peroxidase activity was quenched with hydrogen peroxide reagent for 10 minutes, followed by protease digestion for 30 minutes at room temperature. The following sets of probes were then hybridized for 2 hours at 40^{o}C in a humidity-controlled oven (HybEZ II, ACDbio): Hs-RXFP1 (cat. no. 422821), Hs-FIBP (cat. no. 569781-C2) and Hs-SLC17A7 (cat. no. 415611-C3) to quantify FIBP expression in excitatory (SLC17A7+) layer 5-6 (RXFP1+) neurons; KAZN (cat. no. 569791) and PDGFRA (cat. no. 604481-C3), and HSP90AA1 (cat. no. 477061) to quantify KAZN expression in OPCs (PDGFRA+). Successive addition of amplifiers was performed using the proprietary AMP reagents, and the signal visualized through probe-specific HRP-based detection by tyramide signal amplification with Opal dyes (Opal 520, Opal 570 and Opal 690; Perkin Elmer) diluted 1:300. Slides were then coverslipped with Vectashield mounting medium with DAPI for nuclear staining (Vector Laboratories) and kept at 4^{o}C until imaging.

Imaging and analysis of *in situ* RNA expression

Image acquisitions was performed on a FV1200 laser scanning confocal microscope (FV1200) equipped with a motorized stage. For each experiment and subject, around 10 stack images

were taken to capture at least 20 cells of interest per subject: excitatory neurons (SLC17A7+) from cortical layers 5-6 (RXFP1+), and OPCs (PDGFRA+). Images were taken using a x60 objective (NA = 1.42) with a XY pixel width of 0.3µm and Z spacing of 0.4µm. Laser power and detection parameters were kept consistent between subjects for each set of experiment. Because TSA amplification with Opal dyes yields a high signal to noise ratio, parameters were set so that autofluorescence from lipofuscin and cellular debris was filtered out of the image. Positivity for cell defining markers was determined by bright clustered puncta-like signal present within the nucleus and cytoplasm of the cells. Expression of genes of interest was quantified using the "Analyze Particles" function in Fiji [64]. Stacks were first converted to Z-projections, and for each image cell nuclei of cells of interests were manually contoured based on DAPI expression. Single labeled molecules of RNA were automatically counted in each channel using the find maxima function with a noise tolerance of 350 for FIBP and RXFP1, and 400 for KAZN and PDGFRA. Normalized FIBP and KAZN expression per cell was calculated by dividing FIBP and KAZN raw counts to RXFP1 and PDGFRA raw counts respectively. HSP90AA1 expression was quantified by manually thresholding the signal per image and measuring the percentage of area of the nucleus covered by the resulting mask.

Statistical analysis

No statistical methods were used to predetermine sample size. Sample size was determined based on sample sizes used in previous similar studies. Subjects were assigned to groups based on diagnosis and not by random assignment. All subjects were male, and groups were matched for age (18-87 years), post-mortem interval (12-93 hours), and brain pH (6-7.01). Clinicians were blinded for final psychiatry autopsy diagnosis of MDD case or control. Clustering of single nuclei

gene expression profiles was performed in an unbiased blinded manner. Cluster annotations were assigned after generation of clusters.

Clusters were excluded from downstream analysis if they did not show even contribution from subjects as these clusters are likely to reflect sample specific artifacts rather than biological variability of interest. Single-nuclei were excluded from cell-type clusters based on their level of correlation to the median expression profile of the cluster (lowly correlated nuclei were removed) as detailed above to ensure that differential gene expression analysis was performed using similar nuclei populations from cases and controls. The exclusion criteria were not pre-established and were chosen based on preliminary analysis of the data.

Differential expression analysis between the cases and controls in the snRNA-seq data was performed using linear mixed models implemented in the lme4[57] and lmerTest[58] R packages with biological condition and number of UMIs as fixed effects, the subject and batch as random effects, and a false discovery rate of 0.1 for significance. All DEGs from this analysis were also significantly differentially expressed between cases and controls when re-tested using two-tailed Wilcoxon tests (Supplementary Table 52). For analysis of RNAScope results, two-tailed t-tests were performed with a significance threshold of $p < 0.05$ and data distribution was assumed to be normal but this was not formally tested. For analysis of high-throughput qPCR data two-tailed t-tests or two-tailed Wilcoxon rank sum (i.e. Mann Whitney U tests) were performed, both at a significance threshold of $p <0.05$, and depending on data normality as measured by the Shapiro Wilk's test. The results in Supplementary Tables 43-46 are for genes reliably detected in > 9 subjects per group.

Data Availability

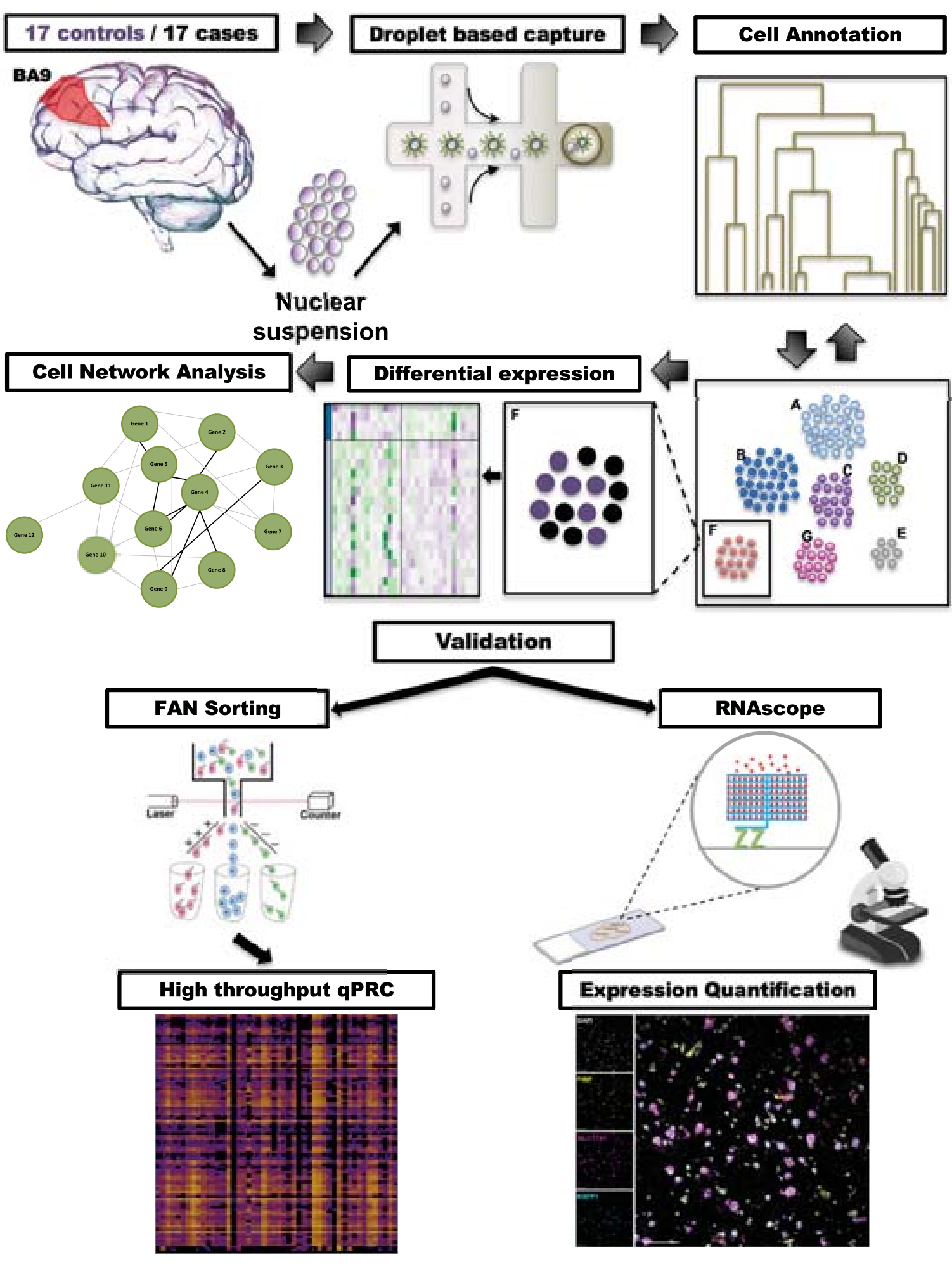
Raw sequencing data, annotated gene-barcode matrix, and lists of cells used for differential gene expression analysis are accessible on GEO using the accession number GSE144136 or using this link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144136. RNAScope and high-throughput qPCR data are available upon request.

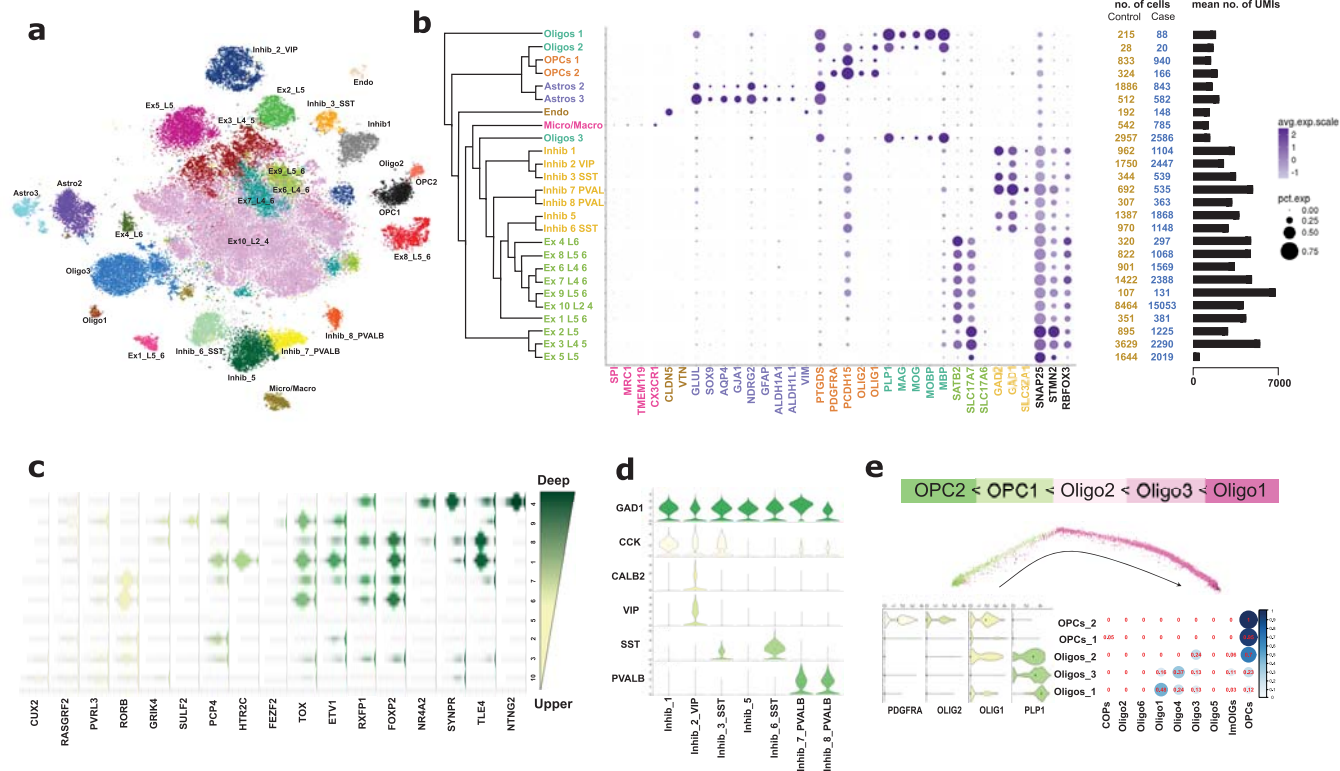
Code Availability

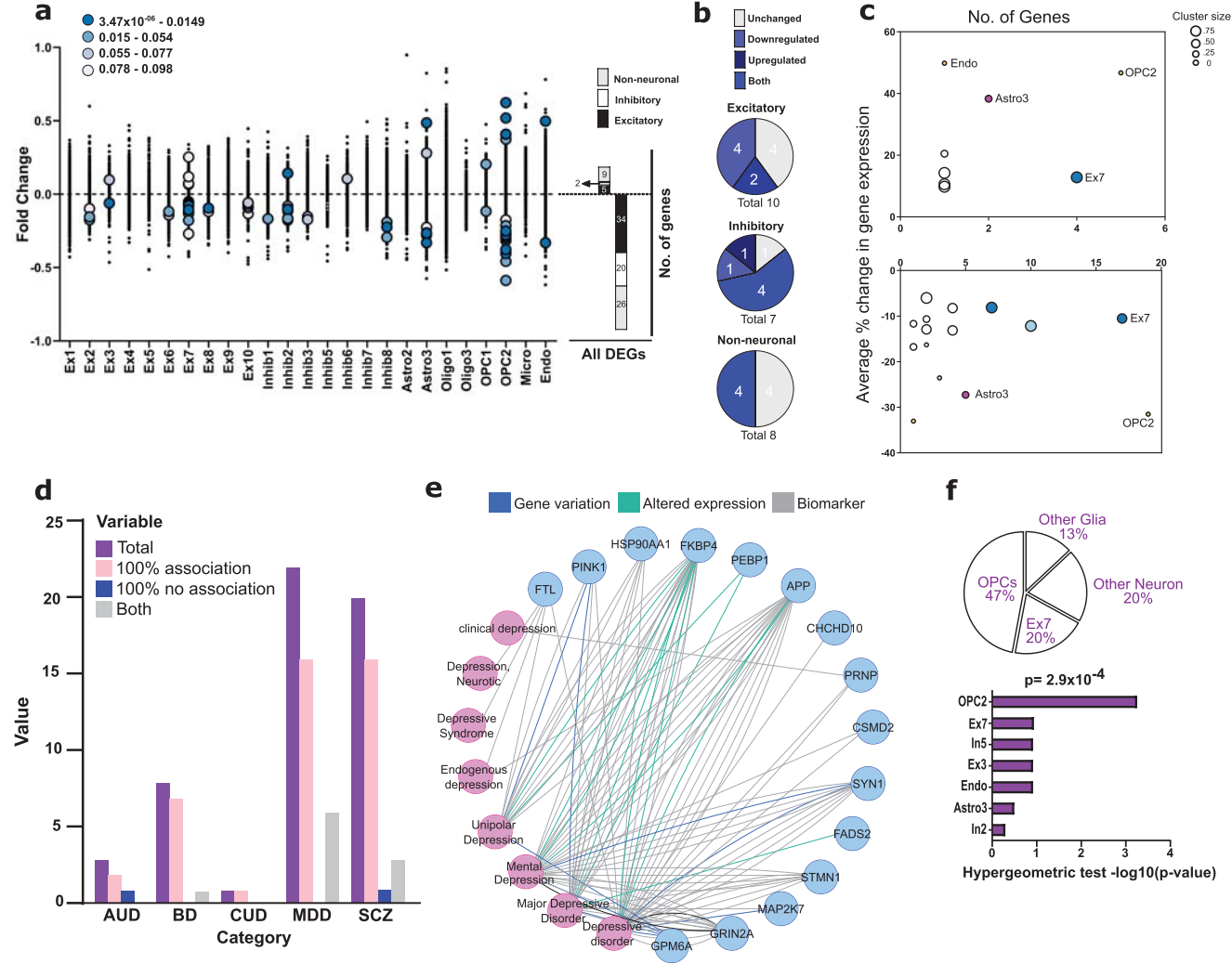
A sample custom R script (Supplementary_R_Script_1.R) used for analyzing high-throughput qPCR data is provided and an R script used to test the statistical significance of CCInx interactions is provided (Supplementary_R_Script_2.R) along with this paper.

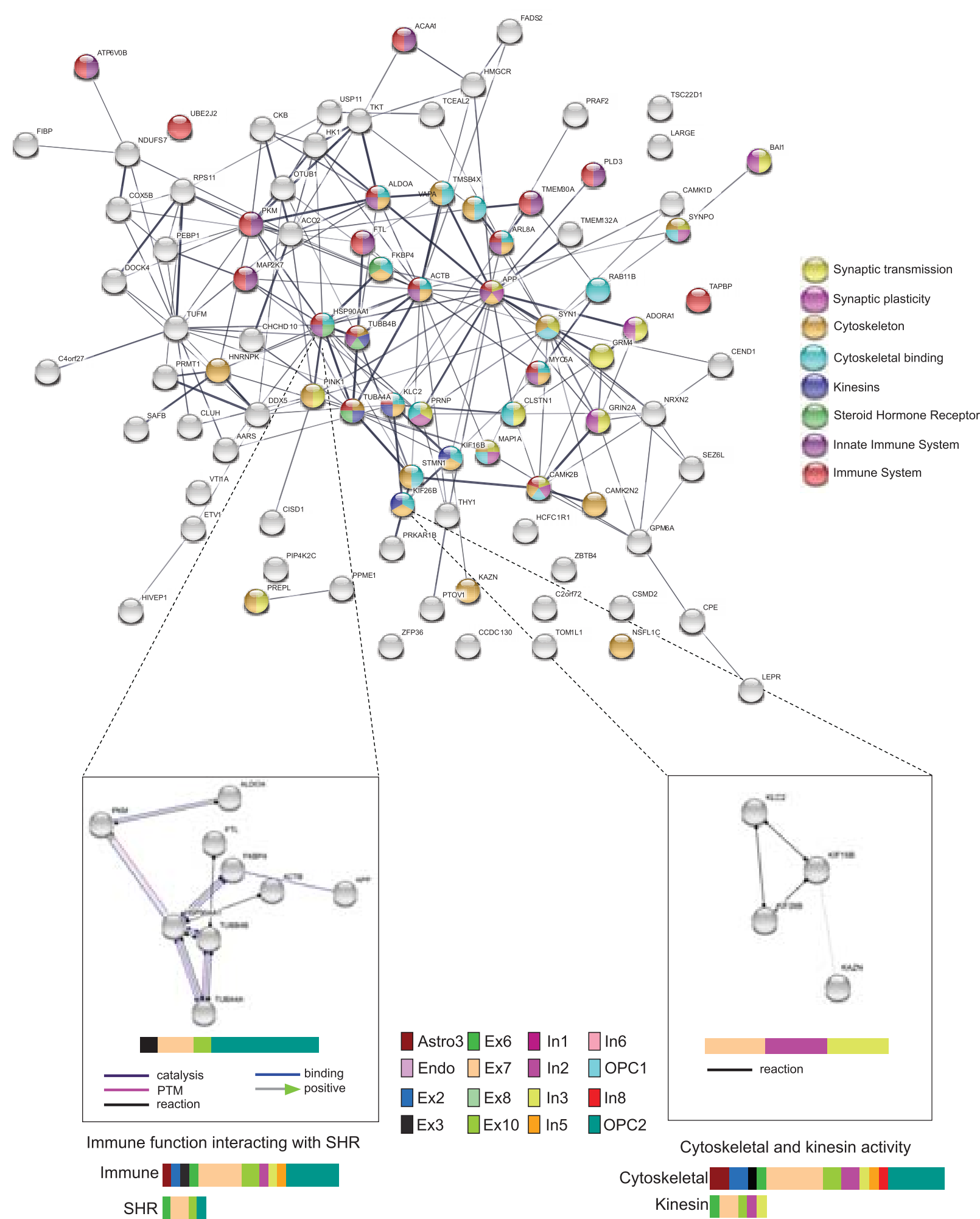
Method-only References

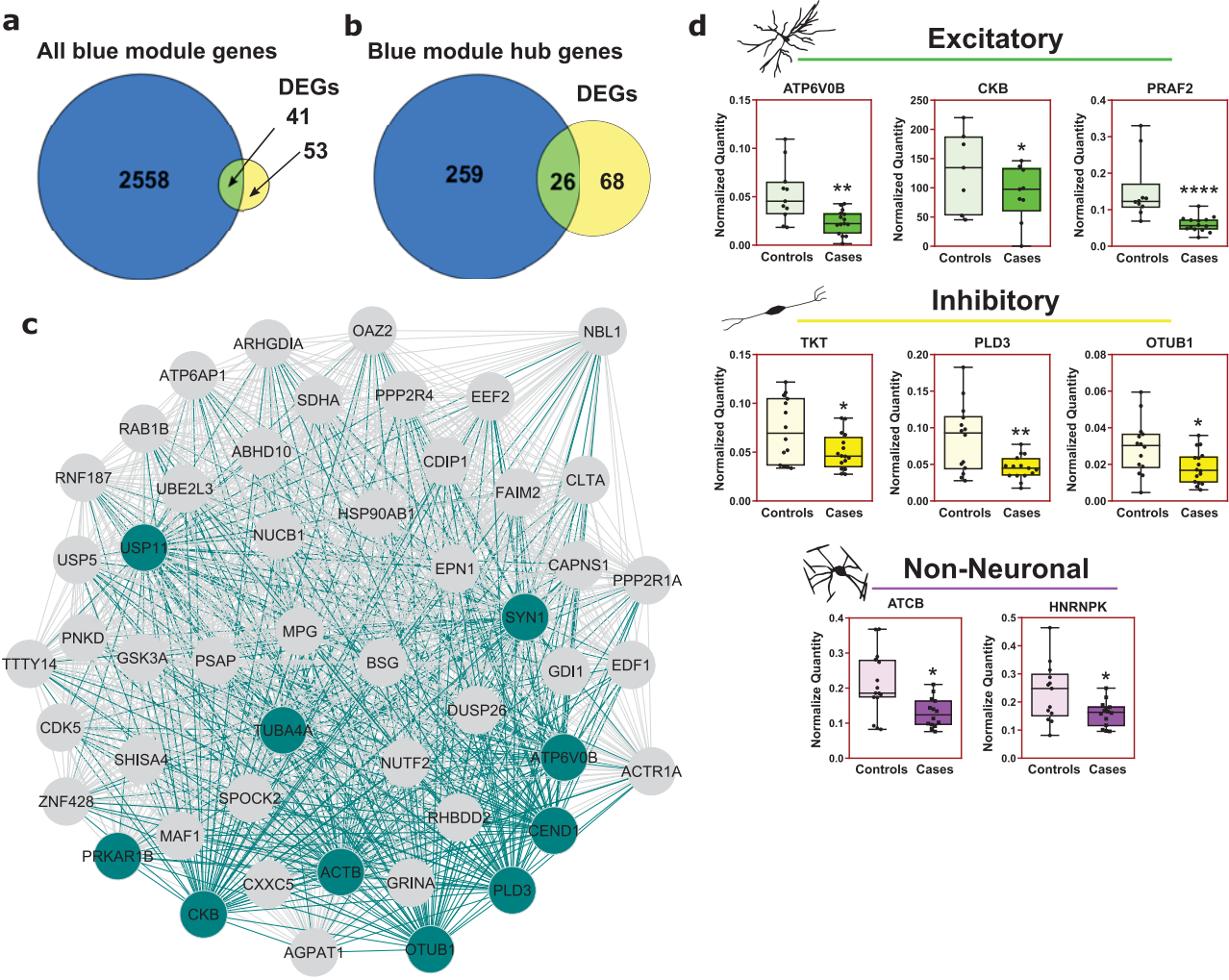
51. Dumais, A.*, et al.* Risk factors for suicide completion in major depression: a case-control study of impulsive and aggressive behaviors in men. *Am J Psychiatry* **162**, 2116-2124 (2005).
52. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2017).
53. Benaglia, T., Chauveau, D., Hunter, D.R. & Young, D.S. mixtools: An R Package for Analyzing Mixture Models. *2009* **32**, 29 (2009).
54. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
55. Hawrylycz, M.J.*, et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012).
56. Lake, B.B.*, et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016).
57. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *2015* **67**, 48 (2015).
58. Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. lmerTest Package: Tests in Linear Mixed Effects Models. *2017* **82**, 26 (2017).
59. Lutz, P.E.*, et al.* Association of a History of Child Abuse With Impaired Myelination in the Anterior Cingulate Cortex: Convergent Epigenetic, Transcriptional, and Morphological Evidence. *Am J Psychiatry* **174**, 1185-1194 (2017).
60. Bayega, A.*, et al.* Transcript Profiling Using Long-Read Sequencing Technologies. in *Gene Expression Analysis: Methods and Protocols* (ed. N. Raghavachari & N. Garcia-Reyero) 121-147 (Springer New York, New York, NY, 2018).
61. Spurgeon, S.L., Jones, R.C. & Ramakrishnan, R. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One* **3**, e1662 (2008).
62. Ximerakis, M.*, et al.* Single-cell transcriptomic profiling of the aging mouse brain. *Nat Neurosci* **22**, 1696-1708 (2019).
63. Gong, T. & Szustakowski, J.D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics* **29**, 1083-1085 (2013).
64. Schindelin, J.*, et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676 (2012).

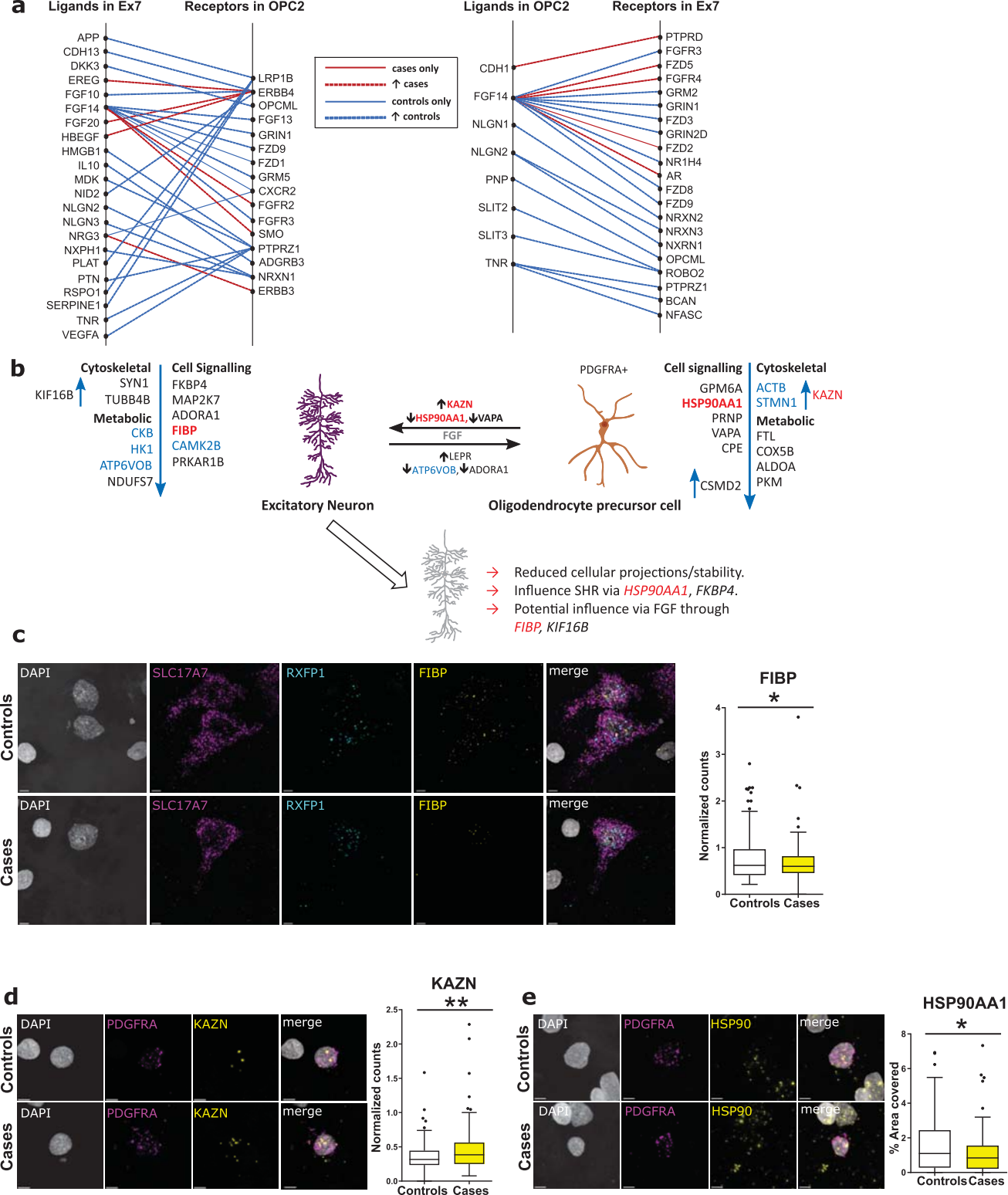

17 controls / 17 cases
Droplet based capture
Cell Annotation
BA9
Nuclear suspension
Cell Network Analysis
Differential expression
Gene 1
Gene 2
Gene 3
Gene 4
Gene 5
Gene 6
Gene 7
Gene 8
Gene 9
Gene 10
Gene 11
Gene 12
A
B
C
D
E
F
G
Validation
FAN Sorting
RNAscope
Laser
Counter
High throughput qPRC
Expression Quantification


a
Inhib_2_VIP
Endo
Ex2_L5
Ex5_L5
Inhib_3_SST
Ex3_L4_5
Inhib1
Oligo2
Ex9_L5_6
OPC2
Astro3
Astro2
Ex6_L4_6
Ex7_L4_6
OPC1
Ex4_L6
Ex10_L2_4
Ex8_L5_6
Oligo3
Oligo1
Inhib_8_PVALB
Ex1_L5_6
Inhib_6_SST
Inhib_7_PVALB
Inhib_5
Micro/Macro
b
Oligos 1
Oligos 2
OPCs 1
OPCs 2
Astros 2
Astros 3
Endo
Micro/Macro
Oligos 3
Inhib 1
Inhib 2 VIP
Inhib 3 SST
Inhib 7 PVAL
Inhib 8 PVAL
Inhib 5
Inhib 6 SST
Ex 4 L6
Ex 8 L5 6
Ex 6 L4 6
Ex 7 L4 6
Ex 9 L5 6
Ex 10 L2 4
Ex 1 L5 6
Ex 2 L5
Ex 3 L4 5
Ex 5 L5
SPI
MRC1
TMEM119
CX3CR1
CLDN5
VTN
GLUL
SOX9
AQP4
GJA1
NDRG2
GFAP
ALDH1A1
ALDH1L1
VIM
PTGDS
PDGFRA
PCDH15
OLIG2
OLIG1
PLP1
MAG
MOG
MOBP
MBP
SATB2
SLC17A7
SLC17A6
GAD2
GAD1
SLC32A1
SNAP25
STMN2
RBFOX3
no. of cells
Control
Case
215 88
28 20
833 940
324 166
1886 843
512 582
192 148
542 785
2957 2586
962 1104
1750 2447
344 539
692 535
307 363
1387 1868
970 1148
320 297
822 1068
901 1569
1422 2388
107 131
8464 15053
351 381
895 1225
3629 2290
1644 2019
mean no. of UMIs
avg.exp.scale
2
1
0
-1
pct.exp
0.00
0.25
0.50
0.75
0
7000
c
Deep
Upper
CUX2
RASGRF2
PVRL3
RORB
GRIK4
SULF2
PCP4
HTR2C
FEZF2
TOX
ETV1
RXFP1
FOXP2
NR4A2
SYNPR
TLE4
NTNG2
d
GAD1
CCK
CALB2
VIP
SST
PVALB
Inhib_1
Inhib_2_VIP
Inhib_3_SST
Inhib_5
Inhib_6_SST
Inhib_7_PVALB
Inhib_8_PVALB
e
OPC2 < OPC1 < Oligo2 < Oligo3 < Oligo1
PDGFRA
OLIG2
OLIG1
PLP1
OPCs_2
OPCs_1
Oligos_2
Oligos_3
Oligos_1
COPs
Oligo2
Oligo6
Oligo1
Oligo4
Oligo3
Oligo5
ImOlGs
OPCs


a
3.47x10-06 - 0.0149
0.015 - 0.054
0.055 - 0.077
0.078 - 0.098
Fold Change
Non-neuronal
Inhibitory
Excitatory
No. of genes
All DEGs
Ex1 Ex2 Ex3 Ex4 Ex5 Ex6 Ex7 Ex8 Ex9 Ex10 Inhib1 Inhib2 Inhib3 Inhib5 Inhib6 Inhib7 Inhib8 Astro2 Astro3 Oligo1 Oligo3 OPC1 OPC2 Micro Endo
b
Unchanged
Downregulated
Upregulated
Both
Excitatory
Total 10
Inhibitory
Total 7
Non-neuronal
Total 8
c
No. of Genes
Cluster size
Average % change in gene expression
Endo
OPC2
Astro3
Ex7
d
Variable
Total
100% association
100% no association
Both
Value
AUD BD CUD MDD SCZ
Category
e
Gene variation
Altered expression
Biomarker
HSP90AA1
FKBP4
PEBP1
PINK1
FTL
APP
clinical depression
CHCHD10
Depression, Neurotic
PRNP
Depressive Syndrome
CSMD2
Endogenous depression
SYN1
Unipolar Depression
FADS2
Mental Depression
STMN1
Major Depressive Disorder
MAP2K7
Depressive disorder
GRIN2A
GPM6A
f
Other Glia 13%
OPCs 47%
Other Neuron 20%
Ex7 20%
p= 2.9x10-4
OPC2
Ex7
In5
Ex3
Endo
Astro3
In2
Hypergeometric test -log10(p-value)


ACAA1
FADS2
ATP6V0B
HMGCR
USP11
TKT
TCEAL2
PRAF2
TSC22D1
UBE2J2
CKB
FIBP
HK1
LARGE
NDUFS7
BAI1
PLD3
RPS11
OTUB1
ALDOA
TMSB4X
TMEM30A
CAMK1D
COX5B
VAPA
PKM
ACO2
TMEM132A
SYNPO
PEBP1
FTL
ARL8A
FKBP4
DOCK4
MAP2K7
ACTB
APP
RAB11B
TAPBP
TUFM
HSP90AA1
SYN1
ADORA1
CHCHD10
TUBB4B
GRM4
C4orf27
HNRNPK
MYO5A
CEND1
PRMT1
PINK1
KLC2
TUBA4A
PRNP
CLSTN1
NRXN2
SAFB
CLUH
DDX5
GRIN2A
AARS
MAP1A
KIF16B
SEZ6L
STMN1
VTI1A
CAMK2B
CAMK2N2
KIF26B
THY1
CISD1
ETV1
HCFC1R1
GPM6A
PRKAR1B
PIP4K2C
ZBTB4
PPME1
KAZN
PREPL
PTOV1
C2orf72
CSMD2
HIVEP1
CPE
ZFP36
CCDC130
TOM1L1
NSFL1C
LEPR
Synaptic transmission
Synaptic plasticity
Cytoskeleton
Cytoskeletal binding
Kinesins
Steroid Hormone Receptor
Innate Immune System
Immune System
catalysis
PTM
reaction
binding
positive
Astro3
Endo
Ex2
Ex3
Ex6
Ex7
Ex8
Ex10
In1
In2
In3
In5
In6
OPC1
In8
OPC2
reaction
Immune function interacting with SHR
Immune
SHR
Cytoskeletal and kinesin activity
Cytoskeletal
Kinesin


a
All blue module genes
2558
DEGs
41
53
b
Blue module hub genes
259
26
68
DEGs
c
ARHGDIA
OAZ2
NBL1
ATP6AP1
SDHA
PPP2R4
EEF2
RAB1B
ABHD10
CDIP1
RNF187
UBE2L3
FAIM2
CLTA
HSP90AB1
NUCB1
USP5
USP11
EPN1
CAPNS1
PPP2R1A
SYN1
PNKD
MPG
TTTY14
GSK3A
PSAP
BSG
GDI1
EDF1
CDK5
TUBA4A
DUSP26
SHISA4
NUTF2
ATP6V0B
ACTR1A
ZNF428
SPOCK2
RHBDD2
CEND1
MAF1
PRKAR1B
ACTB
CXXC5
GRINA
PLD3
CKB
AGPAT1
OTUB1
d
Excitatory
ATP6V0B
CKB
PRAF2
Normalized Quantity
Controls
Cases
**
*
****
Inhibitory
TKT
PLD3
OTUB1
Non-Neuronal
ATCB
HNRNPK
Normalize Quantity


a
Ligands in Ex7
Receptors in OPC2
APP
CDH13
DKK3
EREG
FGF10
FGF14
FGF20
HBEGF
HMGB1
IL10
MDK
NID2
NLGN2
NLGN3
NRG3
NXPH1
PLAT
PTN
RSPO1
SERPINE1
TNR
VEGFA
LRP1B
ERBB4
OPCML
FGF13
GRIN1
FZD9
FZD1
GRM5
CXCR2
FGFR2
FGFR3
SMO
PTPRZ1
ADGRB3
NRXN1
ERBB3
cases only
↑ cases
controls only
↑ controls
Ligands in OPC2
Receptors in Ex7
CDH1
FGF14
NLGN1
NLGN2
PNP
SLIT2
SLIT3
TNR
PTPRD
FGFR3
FZD5
FGFR4
GRM2
GRIN1
FZD3
GRIN2D
FZD2
NR1H4
AR
FZD8
FZD9
NRXN2
NRXN3
NXRN1
OPCML
ROBO2
PTPRZ1
BCAN
NFASC
b
Cytoskeletal
KIF16B
SYN1
TUBB4B
Metabolic
CKB
HK1
ATP6VOB
NDUFS7
Cell Signalling
FKBP4
MAP2K7
ADORA1
FIBP
CAMK2B
PRKAR1B
Excitatory Neuron
↑KAZN
↓HSP90AA1,↓VAPA
FGF
↑LEPR
↓ATP6VOB,↓ADORA1
PDGFRA+
Oligodendrocyte precursor cell
Cell signalling
GPM6A
HSP90AA1
PRNP
VAPA
CPE
CSMD2
Cytoskeletal
ACTB
STMN1
KAZN
Metabolic
FTL
COX5B
ALDOA
PKM
Reduced cellular projections/stability.
Influence SHR via HSP90AA1, FKBP4.
Potential influence via FGF through FIBP, KIF16B
c
Controls
Cases
DAPI
SLIT7A7
RXFP1
FIBP
merge
FIBP
*
Normalized counts
Controls Cases
d
Controls
Cases
DAPI
PDGFRA
KAZN
merge
KAZN
**
Normalized counts
Controls Cases
e
Controls
Cases
DAPI
PDGFRA
HSP90
merge
HSP90AA1
*
% Area covered
Controls Cases